Cognition and Inquiry:

The Pragmatics of Conditional Reasoning

Michael R. Oaksford

PhD

University of Edinburgh

1988

Declaration

I declare that this thesis has been composed by myself and that the research reported therein has been conducted by myself unless otherwise indicated.

Michael R. Oaksford

Edinburgh, 16th October 1988

Acknowledgments

First, I must thank my supervisors Keith Stenning and Robin Cooper. Both have an unerring sense of when support is needed and then providing it. The seeds of most of the ideas for organising an often confusing set of data, were planted by Keith and Robin.

I also owe an enormous debt of gratitude to my friends and colleagues at the Centre for Cognitive Science. They fall roughly into two categories, those who have agreed and those who have disagreed. Both are helpful but the latter perhaps essential. All the following have fitted into both categories at various times: Nick Chater, Mike Malloch, Jon Oberlander, Jerry Seligman. I would also like to thank Robert Dale for his unremitting good humour in the face of my computational illiteracy.

I also thank the Centre's secretaries, Peggy Coonagh and Betty Hughes, for all the support and assistance they have provided over the last three years.

Finally, I thank Becki, Julia and Joanne, without whose love, tolerance and understanding this thesis would not have been written.

This research was supported by a Postgraduate Studentship from the Science and Engineering Research Council.

Abstract

This thesis reports the results of both normative and empirical investigations into human conditional reasoning, i.e. reasoning using *if . . . then* and related constructions.

Previous empirical investigations have concentrated on experimental paradigms like Wason's Selection Task, where subjects must assess evidence relevant to the truth or falsity of a conditional rule. Popperian falsification provided the normative theory by which to assess errorful behaviour on these tasks. However, it is doubtful whether this is an appropriate normative theory from which to derive a competence model of human reasoning abilities.

The relationship between normative theory and competence model need not be direct, no more than the relationship between competence model and performance needs to be. However, research in this area has imported a theory directly into individual psychology from the philosophy of science. On the apparently orthodox assumption of directness, continued adherence to this import may stand in need of re-assessment in the light of the quite radical descriptive inadequacy of falsification as a model of rational scientific inquiry. However, this model also possesses the virtue of relating the interpretation of the rule directly to the normative task strategy.

Hence, this thesis has two aims: first, to retain the virtue of a direct relation between normative task strategy and interpretation while simultaneously offering a competence model which is consistent with more recent and descriptively adequate accounts of the process of scientific inquiry. In Part I, this will involve introducing a semantic theory (situation semantics) and showing that the process of inquiry implicit in this semantic theory is consistent with recent normative conceptions in the philosophy of science.

The second aim is to show that the competence model derived in Part I can provide a sound rational basis for subjects' observed patterns of reasoning in conditional reasoning tasks. In Part II, chapter 5, the data obtained from the Wason Selection Task using only affirmative rules is discussed and the behaviour observed rationally reconstructed in terms of the competence model of Part I. A central concept of that model is *partial interpretation* (motivated by concerns of context sensitivity). *Prima facie* evidence for partial interpretation is provided by the observation of defective truth tables. However, in conditional reasoning experiments using negated constituents, this evidence has been interpreted differently. A subsidiary aim of Part II (which will constitute the largest section of this thesis) therefore concerns the empirical demonstration of the consistency of this data with the competence model.

Contents

Chapter 1: Introduction

This thesis is about human conditional reasoning. The conditional *if...then* construction is central to formal attempts to characterise inferential processes in logic. However, relative to normative logical theories the human data presents a problem. Human conditional reasoning appears beset by various non-logical biases apparently reflecting the influence of content (cf. Wason & Johnson-Laird, 1972; Evans, 1982), memory (Griggs & Cox, 1982), prior beliefs (Pollard & Evans, 1981), resource limitations (Johnson-Laird, 1983) and attentional processes (Evans, 1983a). Thus, normative conceptions of rationality appear radically at odds with people's observed facility for logical thought. This state of affairs has been taken to license the paradoxical conclusion that the only organism apparently capable of formulating systems of pure deductive reasoning, may, after all, be an irrational animal (eg. Stich, 1985).

1.1 Current positions

Within Cognitive Psychology/Cognitive Science this situation has lead to a range of possible responses. First there is a division between *Pragmatic* and *Rationalist* approaches. On the pragmatic view (eg. Evans, 1982), people may have no generalisable unlearned facility for abstract formal reasoning (cf. Evans, 1982). Reasoning is bound to content and limited by selective attention processes (Evans, 1982, 1983a). People are rational, but only in the sense that they are well adapted to their environment. On this view competence models which could provide a rational basis for real human reasoning are unlikely to be found. Subjects can only be classified as falling into error if an appropriate adaptive function for a putative limitation on the cognitive system cannot be identified.

Henceforth, by "competence model" I will mean the following to be understood. A competence model is usually derived from some normative theory, be it semantic theory, probability theory, decision theory or whatever. However, a competence model will also have additional bounds placed on its scope in virtue of common sense assumptions concerning the nature of human cognition. For example, standard logic licenses an infinite list of inferences from any set of premises, none of which would be drawn by any human reasoner (cf. Johnson-Laird, 1986b). Hence, common sense and the known empirical data also enter into the constructive process. In the end, in reasoning research (as opposed to language

production and comprehension) the result should be a model of *rationality*. It should specify the range of inferences one could (reasonably) expect someone to draw, without necessarily specifying anything about how people actually draw them, which is of course the domain of performance.

Rationalists adopt a point of view diametrically opposed to the pragmatist (Henle, 1962; Braine, 1978; Johnson-Laird, 1983, 1986a, 1986b; Cheng & Holyoak, 1985; Cheng et al, 1986). People do possess an ability for abstract reasoning, however there are at least three competing rationalist positions which differ in attitude towards the relationship between *competence* and *performance*. For mental logicians (Henle, 1962; Braine, 1978) formal logics constitute not only the normative theories from which competence models are derived but also, once implemented in the cognitive system, the actual mechanisms of inferential performance. Characteristic errors and biases are put down to errors of interpretation or a lack of a particular inference rule. Thus, error is still possible in a system operating solely on logical principles.

Mental modellers, on the other hand, allow that although normative formal systems are partly constitutive of competence, additional bounds need to be placed on rational performance by various resource limitations (Johnson-Laird, 1983, 1986a, 1986b). This conception (described above) also allows that although competence should be respected by performance, the mechanisms which underlie performance need not be provided by logic *per se* (Johnson-Laird, 1986a:21). Rather, semantic procedures for inference based on mental models are proposed which mirror the observed complexity of human inference patterns. Although logic is not being employed, logically pristine performance is possible. However, limitations on, eg. working memory, will lead to characteristic errors and biases.

A third position holds that although performance is capable of abstract reasoning, this need not be bound by the dictates of any competence model. This is the view implicit in the proposal for the existence of domain specific *pragmatic* reasoning schemas (Cheng & Holyoak, 1985, Cheng et al, 1986). On this view abstract reasoning is specific to particular pragmatic knowledge domains. In contrast both mental logicians and mental modellers propose domain independent mechanisms for inference. On the pragmatic reasoning schemas view, the characteristic errors and biases observed in human reasoning are the result of the inability to access the appropriate schema, or perhaps the elicitation of the wrong schema.

1.2 Points of Departure

The aim of this thesis will be to develop a competence model of the inferential processes involved in conditional reasoning tasks which is adequate both to the data and to certain normative conceptions of rationality. The motivations for this goal came primarily from a dissatisfaction with the assessment that subjects may being doing something irrational on conditional reasoning experiments based on Wason's (1966) selection task. Falsification (Popper, 1959) was taken as normative in this task. However, in the philosophy of science, falsificationism has been recognised as descriptively inadequate for sometime. The competence model to be proposed here will share many of the features of Evans (1982) pragmatic approach. But obviously it is more aligned with the concern of both the mental logicians and modellers, to provide adequate characterisations of what the cognitive system needs to compute as precursor to modelling performance (Johnson-Laird, 1986b). Before indicating how these aims are to be achieved, I will render explicit the grounds for my dissatisfaction with falsification.

The principle experimental paradigm which has fueled speculation concerning the abstract reasoning abilities of human agents has been Wason's selection task (Wason, 1966). In this task subjects must assess evidence relevant to the truth or falsity of a conditional rule. They are presented with four cards and a conditional rule: *if p one side, then q on the other side.* The upwardly turned faces of the four cards correspond to the logical possibilities: p, $\neg p$, q, $\neg q$. In ignorance of the downwardly turned face the subjects' task is to turn only those cards they must in order to determine the rule's truth or falsity. In accordance with the dictates of the normative theory provided by Popperian falsification, which derives directly from the semantics of the material conditional (and the implicit universal quantifier), subjects should turn just the p card and the $\neg q$ card. This is because, logically, these are the only cards with the potential to falsify the rule and, logically, conclusive verification of the rule is impossible (at least in an infinite domain, cf. chapter 4). The empirical finding using abstract material is that subjects tend to turn the p and the q card or the p card only. This response profile accords with a verification strategy whereby subjects are looking only for potentially verifying p, q combinations. Relative to the competence model represented by Popperian falsificationism, this behaviour is irrational.

Falsificationism is a strategy, grounded in formal logic, for the testing of scientific hypotheses, specifically scientific laws. The program which falsification was partly a reaction against was Logical Positivism (Carnap, 1923, 1950; Hempel, 1952, 1965). The logical positivists held a view of meaning derived from Hume and Mach which maintained that the

meaning of an expression was given by its empirical mode of verification. Popper's observations were anti-positivistic in the sense that although one may falsify with certainty, logically one could never verify with similar certainty (but cf. above). However, falsificationism was at least residually positivist in the sense that Popper was still committed to the view that scientific theories and their mode of testing could be formalised. Conditionals and their logic are central to this enterprise since they could provide the means of formally specifying the structure of scientific or causal laws. Understanding the semantic structure of scientific laws is a prerequisite for an understanding of how they can be confirmed. For the logical positivists, a law was simply a universally quantified material conditional which could receive inductive support via the formal definition of a relation of confirmation which held between instances and a law (cf. chapter 3). Popper retains the analysis of laws but replaces confirmation with falsificationism and "corroboration", ie. a law which has withstood our earnest attempts to falsify it, is well "corroborated" (cf. Popper, 1959). A central observation which conditions the position to be adopted in this thesis concerns the fact that within the philosophy of science such attempts to formalise the testing of scientific laws has been rejected as descriptively inadequate for almost 30 years (Toulmin, 1961; Kuhn, 1970, originally 1962; Lakatos, 1970; Putnam, 1974; Feyerabend, 1975).

The philosophy of science is the contemporary locus of discussions of epistemology. Cognitive Science has its origins in "Epistemics" (Goldman, 1986): the study of epistemology which takes the cognitive resources of the knowing agent into account. Contemporary views of the scientific endeavour are particularly sensitive to the historical evidence on the progress of science. And in the historical record it is clear that falsificationist strategies play a very minor and intermittent role in scientific inquiry (Kuhn, 1970; Lakatos, 1970). These issues in the philosophy of science bear on two perennial concerns: scientific realism and rationality. Psychological interest in reasoning research also centres on human rationality and psychology has traditionally drawn on models in the philosophy of science to function as sources of prediction in psychological tasks. This suggests a rather direct relation between a normative theory governing the group activity of science, and a competence model of what individual subjects should do on particular reasoning tasks. I think this orthodoxy is open to question, but not radically. Competence models of rationality are at liberty to modify by common sense, the technical aspects of a normative theory, but nonetheless the central principles governing rational behaviour in both domains remain unchanged. My own view is that mutual influence between these two domains would be the ideal, after all have they have similar goals: to understand how we come to know and reason about our world. However, the traffic has usually been from philosophy of science to

the psychology of reasoning. And in the light of this apparent orthodoxy, if the evidence suggests that certain models in the philosophy of science are descriptively inadequate relative to their own original domain of application, then this might suggest that they should not be taken as normative in conditional reasoning tasks. However, the search for more enlightening competence models in psychology need not abandon the philosophy of science.

Contemporary philosophy of science is a fertile area for the identification of appropriate alternative competence models of conditional reasoning. Moreover, its contemporary concern for descriptive adequacy concerning the processes of inquiry scientists actually employ should provide for psychologically valid competence models. Scientific laws are generalisations involving content. They relate regularly occurring events in the world to other regularly occurring events. It will be shown in chapter 3 that empiricist attempts to reduce the notion of causal law to less problematic concepts fail, causal relations form part of the structure of our world and are not reducible to logical or probabilistic concepts. Knowledge of these relations is fundamental to the ability of organisms to act and react effectively and efficiently to their environment. ie. to act adaptively and therefore rationally. It is these causal and kindred relations which we describe in language using the conditional construction.

This is a familiar view of the semantics of conditionals which finds its origins in Goodman (1983, originally, 1947 & 1955) and Quine (1950). In discussing counterfactual conditionals Quine observes that:

> (1.1) "Any adequate analysis...must consider causal connections, or kindred relationships, between matters spoken of [ie. content] in the antecedent of the conditional and matters spoken of in the consequent" (Quine, 1950:14-15).

In the logic of conditionals it is now a common-place that even the semantics of the indicative conditional is inadequately rendered by the material conditional of standard logic (Nute, 1984; Stalnaker, 1984). Hence, Quine's suggestion can be seen to hold equally for the indicative as well as the subjunctive and contrary-to-fact conditional. In order to adequately characterise the respects in which various conditionals employed in psychological tasks differ, it would be convenient if a semantic theory existed which took Quine's observations and those made concerning scientific laws as the basis for a semantic theory of conditionals. Situation semantics (Barwise & Perry, 1983, Barwise, 1986) adopts just such a position. In situation semantics the locus of meaning is provided by *constraints*, ie. lawlike, causal or explanatory relations which hold between contents. Situation semantics can provide a system of classification which will permit the identification of some important

contrasts between the various conditional rules employed in psychological tasks. The direct connection between situation theory and a contemporary conception of scientific laws in the philosophy of science will also permit the identification of the strategies of confirmation appropriate to conditional rules describing irreducible lawlike relationships or *constraints*.

1.3 Destinations

The principle novelty of the present research inheres in the juxtaposition and application of ideas rather than in any new innovation of my own. The principle semantic theory which supports this research is put to novel use in empirical research. By locating this view in the contemporary literature in the philosophy of science and conditional logics, a model of the inductive process is developed which may have the virtue of psychological plausibility while cohering both with the normative literature and a semantic theory. Five specific applications will prove central to applying this competence model to the empirical results on human conditional reasoning.

First, scientific or causal laws play a central role in predicting and explaining the world. The process of predicting the unknown from the known and thereby gaining information is traditionally classified as *eductive* inference (cf. chapter 4). This serves to contrast the inferential role of the conditional exploited below with its more familiar logical role.

Second, a view of the inductive process will be outlined which is licensed by the need to ground human inferential procedures in actual constraints in the world and which is consistent with recent conceptions in the philosophy of science. This will identify an alternative model to falsification as the rational procedure subjects should adopt in tasks like Wason's selection task.

Third, an important distinction will be drawn between those laws or constraints which apply to instances or single occurrences of events and those which relate discrete events. It will be shown how situation theory provides different interpretations for these cases and how this distinction may lead to subjects' misinterpreting the conditional rules in standard versions of the selection task.

Fourth, it will be shown how negations function differently between these two types of constraint. This will prove central to providing a rational basis for the results obtained in conditional reasoning paradigms where negations are systematically varied between

antecedent and consequent.

Fifth and last, the conception of scientific laws recently emergent in the philosophy of science holds that they are fundamentally context sensitive (Hacking, 1983; Cartwright, 1983). This position is also taken with regard to constraints in situation semantics. One consequence of this view is that interpretation is partial, ie. there can be truth value gaps. This means that false antecedent instances are generally irrelevant to a conditionals truth or falsity. It is a common observation in truth table tasks (eg. Johnson-Laird & Tagart; Evans, 1972) that subjects typically adopt just this interpretation which is described as a "defective truth table". Since this empirical observation is central to the competence theory, most of the empirical work to be reported in this thesis involves re-assessing the import of research on defective truth tables in the light of a competence model which provides a rational basis for the observed behaviour.

1.4 The structure of the thesis

The structure of the Thesis will mirror the distinction between competence and performance. Part I, on competence, will begin in chapter 2 by introducing situation semantics. This will serve to provide a classificatory scheme which can then be employed in later chapters to characterise the important differences between rules used in various psychological tasks. It will also serve to introduce the distinctions which will prove central to providing a rational basis for at least some of the empirical findings on human conditional reasoning.

In chapter 3, the problems raised by empiricism for the conception of objective dependencies in the world will be discussed. This will serve two functions. First to justify the concept of irreducible causal laws which provides the motivation for the situation theoretic concept of a constraint. Second, to identify and resolve further problems which emerge for this semantic theory. This chapter will also include a model of the inductive process licensed by this conception of causal laws. The competence model developed here was equally motivated by consideration of the psychological data.

In the preceding chapters the restrictions of presentation order precluded detailed discussion of the psychological motivation. In chapter 4 this imbalance is redressed. But first the conception of inference licensed by the developing competence model is located in the traditional classification of modes of inference. It is shown how the model provides a rational

basis for some of the data which provided the principle psychological motivation for its development. Two further domains of competence are then looked at. The first relates to some versions of the selection task which employ non-standard procedures and where facilitation of the falsificatory response has often been observed using abstract material. The grounds for this facilitation are located in the specific procedures used which may imply that they are exceptional cases where strategies appropriate to small closed domains are applicable. The implications of partial interpretation and the defective truth table (Wason, 1965) are then discussed. It is argued that the consequences of this empirical observation have been seriously underestimated in the literature.

Part II, on performance, begins in chapter 5 with a discussion of the data on selection tasks which employ only affirmative rules. The results are then provided with a rational basis in the competence model. In this chapter certain process factors need to be introduced to clarify the model's intended import. The model is also given a name: Pragmatic Context Theory. The results obtained when employing thematic content are also introduced and the facilitation of the falsificatory response observed shown to be within the scope of pragmatic context theory.

In chapter 6, the results obtained by Jonathan Evans and his colleagues when systematically varying negations between antecedent and consequent are introduced and critically discussed. The interpretation put on this data (especially Evans, 1983b) is at odds with that provided by pragmatic context theory. Precise hypotheses are therefore formulated which may empirically distinguish between pragmatic context theory and the position taken by Evans (1972, 1979, 1982, 1983b). In chapter 7, the results of these experiments are reported.

In chapter 8, the conclusions licensed by this research are outlined and the consequences for existing theories of conditional reasoning discussed. Some speculations on the nature of the general cognitive architecture which would be consistent with the conclusions of this research are also offered.

Part I:

Competence

Chapter 2: Situation Semantics

2.1 Introduction

There are two principle reasons for introducing situations semantics. First, there is a need for a system of classification in order to characterise the semantic/pragmatic content of conditionals. This is required to identify the respects in which the semantics of conditional rules used in psychological experiments differ. Second, the reasons for introducing situation semantics in particular are tied to the preliminary outline presented in chapter 1, of the kind of competence theory psychology requires. Following Quine (1950), it was suggested that the kind of theory psychology needs must take subject matter or content into account. This also accords with the preliminary observations made in chapter 1 concerning the evidence on the psychological importance of content in human reasoning experiments. Another aspect of those results which was highlighted in chapter 1, concerned the apparent partiality of subjects' interpretations of conditionals. This functions as a further motivating factor in introducing situation semantics. Content and partiality are central to the situation theoretic perspective.

2.2 Situation Semantics

Semantics is about assigning bits of the world to sentences. In model theoretic semantics this is done by defining an abstract model of the world (this may include bits of the mind, cf. intensional logics) which can be systematically assigned to parts of a language via a recursive definition which defines a truth predicate for that language. The world is modeled using the resources of set theory. However, the objects and relations which can be defined in set theoretic terms may provide to coarse grained an account to get at subject matter (Barwise, 1986). Hence, a prior requirement is to develop an account of the contents of the world which can subsequently be used to provide a semantics for natural language. This is the enterprise of *situation theory*. The current stage of development is informal in the sense that there is no mathematical model of situation theory. As it stands, situation theory is a formalism for talking about the world from the situation theoretic perspective. It is this formalism which will be introduced in this chapter.

First, by way of further motivation, some observations will made on how situation semantics views partiality and content.

2.2.1 Partial interpretation

Partial interpretation was introduced in the last chapter via the observation that in many experiments there is evidence that people take some conditions to be irrelevant to the truth or falsity of a conditional. This phenomenon can be captured logically by allowing a third truth value and redefining the truth tables for the connectives accordingly (cf. Haack, 1975). Alternatively, bivalence can be retained but truth relativised to knowledge or information states (Kripke, 1965). This is the option taken by Kripke (1965) in providing a semantics for intuitionistic logic. The idea has been taken up again recently by Veltman (1985, 1986). An information model (Veltman, 1985, 1986) contains a partially ordered set of information states each consisting of a set of atomic propositions. A partial valuation function assigns truth values to these propositions. "Partial" here means that for some propositions the valuation function, which takes propositions as arguments, is undefined. Hence, partiality is captured while retaining bivalence. The recursive truth definition for the connectives is defined relative to information states and specifies conditions on the valuation function which must obtain for the connective to be true or false (both must be defined since although there are only two truth values the valuation function is undefined for some propositions).

In Veltman's system truth is relativised to an epistemic concept, ie. truth on the basis of the evidence. We can contrast the situation theoretic concept of a *situation* with the concept of an information state. An information state is a psychological concept. Whereas, a situation is a local part of the world. A situation can be viewed as an individuated fragment of the world to which an agent may attend at any given time. The limits on a situation are given by an agents perceptual/attentional resources. So, a situation is somewhere in between a purely psychological something and reality. What someone attends to in a situation may well depend on prior knowledge but the situation is not defined in terms of what that person knows. More will be said on the status of situation theoretic objects in the next chapter. For the moment, situations can be contrasted with information states insofar as the former but not the latter are taken to be real parts of the world.

2.2.2 Information conditions

A major point of contrast is that rather than unstructured atomic propositions, the situation theoretic equivalent to the valuation function is defined over structured objects which describe the contents of a statement. This marks a shift in emphasis away from truth conditions and towards *information conditions*. Situation theory is attempting to provide an account of *informational content*, it, therefore, wants to provide an account of the conditions under which a statement is informational. This contrasts with having propositional content as the primary goal. Here the concern is to provide an account of the conditions under which a statement is true. The difference is that although if a statement is informational then it is true, it is not the case that if it is true, then it is informational.

This can be seen via the additional observation that truth conditions are relative to language or some other representational scheme, eg. a representational mental state. However, information is representation independent. Smoke carries the information that fire whether anyone represents the fact or describes it in language. Because of this real world *relationship* an utterance "smoke" can carry the information that there is fire. If it does carry this information on a particular occasion of use, then "smoke" is true. Whether it does or not is determined by the structure of the world. And whether it is informational for any individual is determined by whether they are attuned to that structure. This is the principle motivation for proposing a semantic theory which talks about content. The relations which license information flow hold between particular parts of the world.

Of themselves, these twin motivations make situation semantics of psychological interest. Both factors, partiality and content dependence, are common observations in the psychological data (for a complete justification of this claim cf. chapter 4). The formalism which attempts to capture these properties will now be introduced.

2.3 Situation theory

The following account of situation theory is based on unpublished manuscripts by Barwise (1987) and Israel & Perry (1987). The reason for not using a standard text, for example Barwise & Perry (1983) or Barwise (1986), is due to the state of flux in which situation semantics currently finds itself. The need to fully articulate situation theory as a precursor to a full blown situation semantics is a relatively recent realisation. Moreover, the richness of the more recent, but admittedly preliminary, formulations will be required later on. It

would be a marked step forward in providing an empirically adequate semantic theory if the psychological applications to be discussed were in turn to motivate aspects of the theory itself. With this in mind, possible, but minor, extensions to the theory will be proposed. These extensions will all be based on the need to characterise some aspect of the semantics/pragmatics of conditional sentences as they appear to influence the psychological data.

2.3.1 Situations

Situation Theory is committed to the existence of a concrete, structured reality. Reality has parts, called situations but does not include possible ways the world may have been. However, reality can include mental states corresponding to ways people believe the world to be. But that people possess these mental states is as much a fact about the world as the existence of mind independent physical states. Situation Theory is not committed to there being a largest total situation of which all others are parts, it is a *localist* theory (cf. chapter. 3).

2.3.2 Relations and states of affairs

In order to analyse reality some system of *classification* is required which for the purposes of semantic theory reflects how that reality is individuated. This will include domains of situations $\{s, s' ...\}$, relations $\{R, R' ...\}$, spatio-temporal locations $\{l, l' ...\}$ and individuals $\{a, a' ...\}$.

A relation R comes with a set of *argument roles*. The relation of *traveling*, for example, comes with the roles of, *traveler*, *mode of transport ("mode")*, *destination* and *location of traveling*. Locations may be spatially and/or temporally extended.

Individuals or objects must be of the appropriate sort to play these roles. The traveler must be an animate object, the mode is obvious, ie. bike, car, plane etc., destination a physical location.

Given an appropriate assignment of objects to its roles, a relation gives rise to an *issue* concerning whether or not the objects stand in that relation. There are two possibilities, each called a *state of affairs* (soa).

For example, take the relation of traveling, let Johnny be the traveler, train be the mode, Manchester be the destination and the 20th April 1988 be the location (henceforth: *l*), then the following two soas arise:

(2.2) *<<Traveling, l: 20-4-88, traveler: Johnny, mode: train, destination: Manchester; 1>>*, and

(2.3) *<<Traveling, l: 20-4-88, traveler: Johnny, mode: train, destination: Manchester; 0>>*

(2.2) resolves the issue positively, whereas (2.3) resolves it negatively; (2.2) has positive *polarity*, and (2.3) negative polarity. Normally argument roles are suppressed and, henceforth, they will not be included in descriptions of soas. The objects assigned to the various roles are called the *minor constituents* and the *traveling* relation in this soa is the *major constituent*. soas are the structured objects which enables situation theory to characterise content.

2.3.3 Facts and Partiality

A fundamental relation in situation theory is "⊨", the *holds-in* relation. If an soa σ holds in a situation *s*, ie.

(2.4) $s \models \sigma$

then σ is a *fact*. If σ's dual (same assignment of constituents but different polarity) holds in *s*, then this can be expressed as:

(2.5) $s \models \neg\sigma$

Situations are only partial bits of the world. Therefore, for any given issue a situation *s* may resolve σ positively, or negatively, *or* *s* may fail to resolve the issue one way or another. The converse relation "⊭" for "does not hold in" expresses this. If neither σ nor its dual ¬σ holds in *s*, this can be expressed:

(2.6) $s \not\models \sigma,\ s \not\models \neg\sigma$

This aspect of "⊨" mirrors the partial valuation function for Veltman's information models previously discussed. The general form of a proposition is a pair: $s \models \sigma$, which can be true, false or undecided.

Since, situations are only partial bits of the world, one situation *s* can be part of another

situation s'. Hence, another fundamental relation is "<", the *part-of* relation.

(2.7) $s < s'$

This is the case iff:

(2.8) If $s \models \sigma$, then $s' \models \sigma$.

However, there may be no situation s such that all other situations are part of it, ie. there is no "maximal" situation. This is equivalent to the partial ordering over information states in Veltman's information models. The important difference is that, whereas a chain of information states always ends in a complete, "maximal", information state, this is not the case for situations.

2.3.4 Saturated, unsaturated and parametric soas

In a *saturated* soa there is an appropriate *assignment* of some object to every argument role, as in (2.2) and (2.3) above. Thus, (2.2) would be the interpretation of the assertion that:

(2.9) Johnny traveled to Manchester by train.

In an *unsaturated* soa one (or more) argument role(s) is not assigned an object, for example:

σ_u: *<<Traveling, l, Johnny, ??, train; 1>>*

Here it still makes sense to say that there is a situation s, such that σ_u holds in s, ie. $s \models \sigma_u$. That assignments can be partial expresses the fact that sometimes it is not possible, or it may not be important to fully describe a *soa*, it nonetheless raises an issue which can be resolved. So, σ_u would be the interpretation of the assertion that:

(2.10) Johnny traveled by train.

This situation relates to the two relations above in an intuitive way. For any situation s, $s \models \sigma_u$ iff there is some s' such that $s < s'$ and $s' \models \sigma_u$. So an unsaturated soa holds in situation s iff s can be incrementally extended to a situation s' such s is a part of s', and in s' the *destination* role is filled in some appropriate way.

However, sometimes although precisely where Johnny traveled may not be important that he traveled *somewhere* is. In this case a *parameter* must be assigned to the argument role:

(2.11) *<<Traveling, l, Johnny,* **x**, *train; 1>>*

So, a domain of parameters: **x**, **y**, **z**...is required (these function rather like free variables); these are enboldened to distinguish them from individual constants: *a*, *b*, *c*..., and possibly bound variables x, y, z... (2.11) would be the interpretation of the assertion that:

(2.12) Johnny traveled somewhere by train.

Parametric soas ("**psoas**") are denoted "Σ".

2.3.5 Appropriate assignments, anchors and proper anchors

Argument roles place *restrictions* on assignments such that they are appropriate. For example, the *mode* argument role restricts assignments to modes of transport, so *train, car...etc.* are appropriate assignments whereas, Manchester, table, the number 77...etc. are not.

It does not make sense to ask whether a **psoa** holds in a given situation. Nonetheless parametric claims can be made about a situation, eg. (2.12). However, to determine whether:

(2.13) $s \models$ *<<Traveling, l, Johnny,* **x**, *train; 1>>*,

requires **x** to be filled. Parametric argument roles are filled using *anchors*; possibly partial functions from parameters to objects. Anchors are to parameters what assignments are to argument roles. Let Σ(destination: **x**) denote the **psoa** in (2.11). That an anchor *f* anchors this **psoa** can be written Σ(destination: **x**)[*f*]. If *f* : **x** |-> Manchester, then *f* is said to be *proper* as Manchester is an appropriate sort of thing to fill this role.

2.3.6 Situation types

Psoas can be used to generate *situation types*. For example, (2.11) is a **psoa** Σ:

(2.14) Σ: *<<Traveling, l, Johnny,* **x**, *train; 1>>*

A type T, where the generating type is given by Σ, is expressed as T_Σ, and when a situation *s* is of type T_Σ, this is written $s{:}T_\Sigma$. A situation is of a given type T, if there is an anchor for the parameters of the generating **psoa** Σ (2.14), such that $s \models \Sigma[f]$.

It is not necessary to generate types only from **psoas**. To preserve generality, it can be allowed that non-parametric **soas** can also generate types, eg.

(2.15) σ: *<<Traveling, Johnny, Manchester, train; 1>>*

These can be seen as a special case. Henceforth, only parametric types will be discussed, so the unmarked term "type" will be restricted to parametric types. Later on, if a non-parametric type is introduced, it will be explicitly marked.

This provides a dimension of *generality* versus *specificity* on types. Types generated from non-parametric **soas** are clearly the most specific. All argument roles have specific constituents assigned. A type generated from a **psoa** whose argument roles are all assigned parameters is clearly the most general type. There is a dimension of variation between these two extremes dependent on the number of argument roles assigned parameters.

2.3.7 Restrictions

Restrictions can be placed on anchors such that only anchors which also satisfy the restriction will be admitted as proper anchors for restricted parameters. So for example, in (2.11), it could be further specified that **x** be restricted to British travel destinations, ie. any anchor for **x** must also provide an anchor for:

(2.16) *<<British_travel_destination*, **x**; *1>>*

This concept of a restriction will prove important later on in defining *taxonomic* constraints.

2.3.8 Constraints

The class of relations which produce *constraints* are central to the situation semantic account of meaning and information. Constraints control the flow of information within a situation. Being *attuned* (cf. below) to a constraint allows people to draw inferences to gain information about their world. Constraints are based on *involves* relations. There are many involves relations, some of the important ones are as follows:

[l]=>: INVOLVES-LOGICALLY
[a]=>: INVOLVES-ANALYTICALLY
[n]=>: INVOLVES-NOMICALY
[c]=>: INVOLVES-CONVENTIONALLY

These relations take types as arguments. A constraint is a soa based on an involves relation and having positive polarity. For example, (2.17) is a ubiquitous constraint.

(2.17) $\ll[i]\Rightarrow, T_{\Sigma}, T'_{\Sigma'}; 1\gg$

If a situation s supports (2.17), then $s{:}T_{\Sigma}$ involves s being of the consequent type $s{:}T'_{\Sigma'}$. That is, any anchor f for the parameters of T_{Σ} can be extended to an anchor g which also anchors the parameters of $T'_{\Sigma'}$.

The following example of an analytic constraint instantiates the schema in (2.17), let:

Σ: $\ll$*Traveling*, l, x, y, z; *1*$\gg$, and

Σ': $\ll$*Mode_of_transport*, l, x; *1*$\gg$

Then the following is an analytic constraint:

(2.18) $\ll[a]\Rightarrow, T_{\Sigma}, T'_{\Sigma'}; 1\gg$

This constraint is unconditional. Any traveling event will involve the traveler employing a mode of transport, be it his legs, a train, bus etc. There are no other conditions which need to be fixed before inferring that a mode of transport is being employed if a traveling event is taking place. However, most constraints are *conditional*, a situation s supports such a constraint only if s supports some other types.

For example, if an object is unsupported it falls, expresses the constraint that being unsupported involves falling. This can be expressed as the following nomic constraint:

Σ: $\ll$*Unsupported*, l, x; *1*$\gg$, and

Σ': $\ll$*Falling*, l, x; *1*$\gg$

Then the following is a nomic constraint:

(2.19) $\ll[n]\Rightarrow, T_{\Sigma}, T'_{\Sigma'}; 1\gg$

However, this constraint can only be supported in a situation where there is zero gravity. So, all situations which support (2.19) must be of type $T''_{\Sigma''}$, where:

Σ'': $\ll$*Zero_gravity*, l; 0$\gg$

This can be expressed by allowing the general form of a constraint to be a ternary relation:

(2.19') $\langle\langle[i]\Rightarrow, T_{\Sigma}, T'_{\Sigma'}, T''_{\Sigma''}; 1\rangle\rangle$

T_{Σ} is called the *indicating* type, $T'_{\Sigma'}$ the *indicated* type and $T''_{\Sigma''}$ the background or connecting type.

The background type can be complex. When a situation supports a constraint many other background types will also be fixed, ultimately their precise nature may be inefable. This captures the conditional nature of most constraints. Some constraints are ubiquitous in so far as they hold in all situations. Perhaps some high level physical laws are ubiquitous in this sense. Of the conditional constraints where one or more background type needs to be fixed, some are more conditional than others. The more background types which would need to be fixed the more conditional it is. The ubiquity or conditionality of a constraint should not be confused with the generality or specificity of the related types.

The arguments of a constraint are usually quite general types. In a particular situation the parameters of these types will be anchored to specific objects. Capturing this is why the background type was also described as the *connecting type*. Additional more specific types may be included in the background or connecting type which explicitly place restrictions on the anchors for the indicating and indicated types. These types *connect* the constraint to a specific instantiation. So, for (2.19), $T''_{\Sigma''}$ may include:

$T''_{\Sigma''}$: *<<Johnny's ball* **x**; *1>>*

This restricts any anchor *f* for the indicating and indicated type such that *f*: **x** |-> *Johnny's ball.*

It is not always the case that constraints are defined over the most general types. For example, suppose you are disposed to buy a newspaper on your way to work. This can be expressed as the following dispositional constraint:

Σ: *<<Going_to_work*, **l**, **you**; *1>>*, and

Σ': *<<Buying*, **l**, **you**, **x**; *1>>* & *<<newspaper*, **l**, **x**; *1>>*

Then the following is a dispositional constraint:

(2.20) $\langle\langle[d]\Rightarrow, T_{\Sigma}, T'_{\Sigma'}; 1\rangle\rangle$

This constraint is specific in so far as it cannot be expressed as the instantiation of any more general constraint. When this is the case it is inappropriate to describe it more generally, and then add further connecting types. The higher level generality is misleading:

going to work does not involve buying a newspaper, unless *you* are disposed to do so.

A special case is where the parameters of the indicated type are included in the parameters of the indicating type. For example, someone, say Johnny, may be disposed to travel to Manchester only by train. This can be described by the following constraint:

T_Σ: <<*Traveling*, **l**, *Johnny*, *Manchester*, **z**; *1*>>, and

T'_Σ: <<*train*, **z**; *1*>>

C: <<[d]=>, T_Σ, T'_Σ; *1*>>

Here the parameter **z** in T′ is included in the parameters of T. When a situation *s* |= C, then if *s*:T[*f*], then *s*:T′[*f*]. It is the *same* anchor *f* which must anchor both indicating and indicated types, this contrasts with (2.20) where there are additional parameters in the indicated type not included in the indicating type. In C the indicated type adds *restrictions* on appropriate anchors for unassigned minor constituents in the indicating type.

Constraints of this form will be called *taxonomic*. This is because they induce a restriction on an anchor *f* to a token *a* such that if *a* is of the indicating type then *a* is also of the indicated type. Such restrictions give rise to a class inclusion hierarchy in the obvious way. Taxonomic constraints do not form a separate class: the constraint C is dispositional, but it is of taxonomic type. To mark this the nomenclature "[d_t]=>" will be used.

Taxonomic constraints serve to encode the distinction between involves relations which ascribe properties to instances of objects or single occurrences of events rather than relating discrete occurrences. C describes single occurrences of traveling events, which are restricted such that the mode parameter can only be anchored to tokens of the type train. In non-taxonomic constraints discrete events are related, by some higher order involves relation. This distinction will prove central in providing a rational basis for peoples' behaviour on various conditional reasoning tasks in subsequent chapters.

Constraints are higher level relations. Along with the involves relations already introduced there may be other slightly lower level kinds. For example, within the conventional constraints a permission involves relation will be found. Qua relation, constraints also impose restrictions on proper assignments to the argument roles which can be filled by types. For example, the permission relation will demand that the indicating type be an *action* and its indicated type a *precondition*. So, for example, that to enter the Phillipines involves having been inoculated against cholera may be expressed as follows, making the argument roles

explicit:

T_Σ: <<*Enter*, *Phillipines*, **x**; *1*>>, and

$T'_{\Sigma'}$: <<*Inoculated*, *Phillipines*, **x**, *cholera*; *1*>>

(2.22) <<[pm]=>, *action*: T_Σ, *precondition*: $T'_{\Sigma'}$; *1*>>

Constraints are the situation theoretic objects which function as the interpretations of conditionals in situation semantics. In general conditional sentences are interpreted as proposing the existence of a constraint. However, there may be no 1 - 1 mapping between constraints and conditionals sentences. This can be understood by analogy with the inference rules which logic attaches to the conditional construction. Different inference rules will apply dependent on the subjunctive or indicative mood of the conditional, or perhaps in tense logic on the temporal relationship which holds between antecedent and consequent. By anology, dependent on what contextual information is available either directly in the environment or indirectly in an agents prior beliefs, a conditional may propose different or multiple constraints to be operative within a given situation. Exampling this phenomenon will be postponed to later chapters, where it will prove central to explicating the rational basis of peoples' inferential behaviour in conditional reasoning tasks.

2.3.9 "=|": The precludes relation

The precludes relation is a primitive relation in situation theory, although it is fundamentally related to the involves relations. Preclusion also takes types as arguments in exactly the same way as the involves relations.

(2.25) <<[*i*]=|, T_Σ, $T'_{\Sigma'}$; *1*>>

If a situation *s* supports (2.25), then *s*:T_Σ precludes *s* being of the consequent type *s*:$T'_{\Sigma'}$. That is, any anchor *f* for the parameters of T_Σ can not be extended to an anchor *g* which also anchors the parameters of $T'_{\Sigma'}$.

The preclusion relation captures the fact that many events may be related to each other but in a negative way. In the causal case, events may be positively or negatively causally related. For example, smoking causes heart disease, and exercise prevents heart disease. As the involves relations concern the positive relations between events, so the precludes

relations concern the negative relations. The rest of this chapter is not concerned with situation theory per se, but rather using it to characterise some distinctions which will prove important in subsequent chapters. We will begin by looking at how negative soas function to identify contrast classes.

Take a previous example:

(2.26) *<<Traveling*, **l**, *Johnny, Manchester, car; 0>>*

What is the cognitive significance of assigning negative polarity to this soa? The force it possesses depends on which constituent is *focused* upon. In natural language negative focus is usually marked by intonation. So let us look at the way this works in natural language. The stressed word is enboldened

(2.27) Johnny didn't travel to Manchester by **car.**

(2.27′) Johnny didn't travel to **Manchester** by car.

(2.27″) **Johnny** didn't travel to Manchester by car.

(2.27‴) Johnny didn't **travel** to Manchester by car.

In each of (2.27) - (2.27‴) intonation serves to identify which minor constituent is being denied. Each constituent is an appropriate assignment to some argument role. And the argument roles serve to identify the relevant contrast class. In (2.27), it is denied that Johnny traveled to Manchester by *car*, so the relevant contrast class is modes of transport, ie. Johnny traveled to Manchester by some mode of transport other than the car. In (2.27′), it is denied that Johnny traveled to *Manchester* by car, so the relevant contrast class is travel destinations, ie. Johnny traveled by car to some destination other than Manchester. In (2.27″), it is denied that *Johnny* traveled to Manchester by car, the relevant contrast class is traveler, ie. someone else other than Johnny traveled to Manchester by car. (2.27‴) indicates that the whole traveling event did not occur. For (2.27) - (2.27″), focus can be encoded by placing a negative restriction on anchors for a particular constituent. For example, take (2.27):

(2.26) *<<Traveling*, **l**, *Johnny, Manchester*, **x**; *1>>* & *<<Car*, **l**, **x**; *0>>*

For *n*-ary relations, if the soa has negative polarity, then it can be assumed that the issue raised concerns whether a whole event defined by the major constituent held in a particular situation.

The relations which act as major constituents of soas may function differently with respect

to their domains of complementation. For example, (2.29)

(2.29) *<<Mode*, l, x; *1>>* & *<<Car*, l, x; *0>>*

identifies a full contrast class consisting of modes of transport other than cars. This will include sub types: train, aeroplane etc. of which individual trains, aeroplanes etc. are tokens. However, many relations may not identify such extensive contrast classes but rather may define other gradations of a continuum. For example:

(2.30) *<<Zero_gravity*, l; *0>>*

serves to identify a soa where gravity has some positive value. Similarly, the dimension of variation may be dichotomous:

(2.31) *<<Stationary*, l, x; *0>>*

identifies a positive soa where x is in motion. These *antonymic* cases may be contextually defined. For example:

(2.32) *<<Serves-drinks*, l, *Mary*, x; *1>>* & *<<Tea*, l, x; *0>>*

The dimension of variation is drinks, but in a context where Mary serves tea she is unlikely to serve cocktails, so perhaps coffee is the most likely contextually defined antonym.

Preclusion serves to identify why various background types need to be specified for a constraint to hold in a situation. If the duals of any background type held in a situation, then they would preclude the indicating type of the constraint from holding. Thus, for a situation to support the constraint in (21) concerning unsupported objects falling, it must also support the condition in the background type that gravity is non-zero. This is because if gravity were zero, this would preclude an object from falling.

2.4 Inference and Information Gain.

"Inference" in situation theory concerns the constraints between mental states which permit those states to track reality. Diagram 1, taken from Barwise (1984), serves to illustrate this: There are constraints which hold between various mental states and states of the world. Types of mental state are denoted "ΨT". So a constraint which holds between mental types and types in the world is given by C: ΨT => T. If an agent *a* is in mental state Ψσ which is of type ΨT, then this is normally because there is an actual state of affairs σ which is of type T. In Barwise (1984), for *a* to be attuned to a constraint C′: T => T′ amounts to the

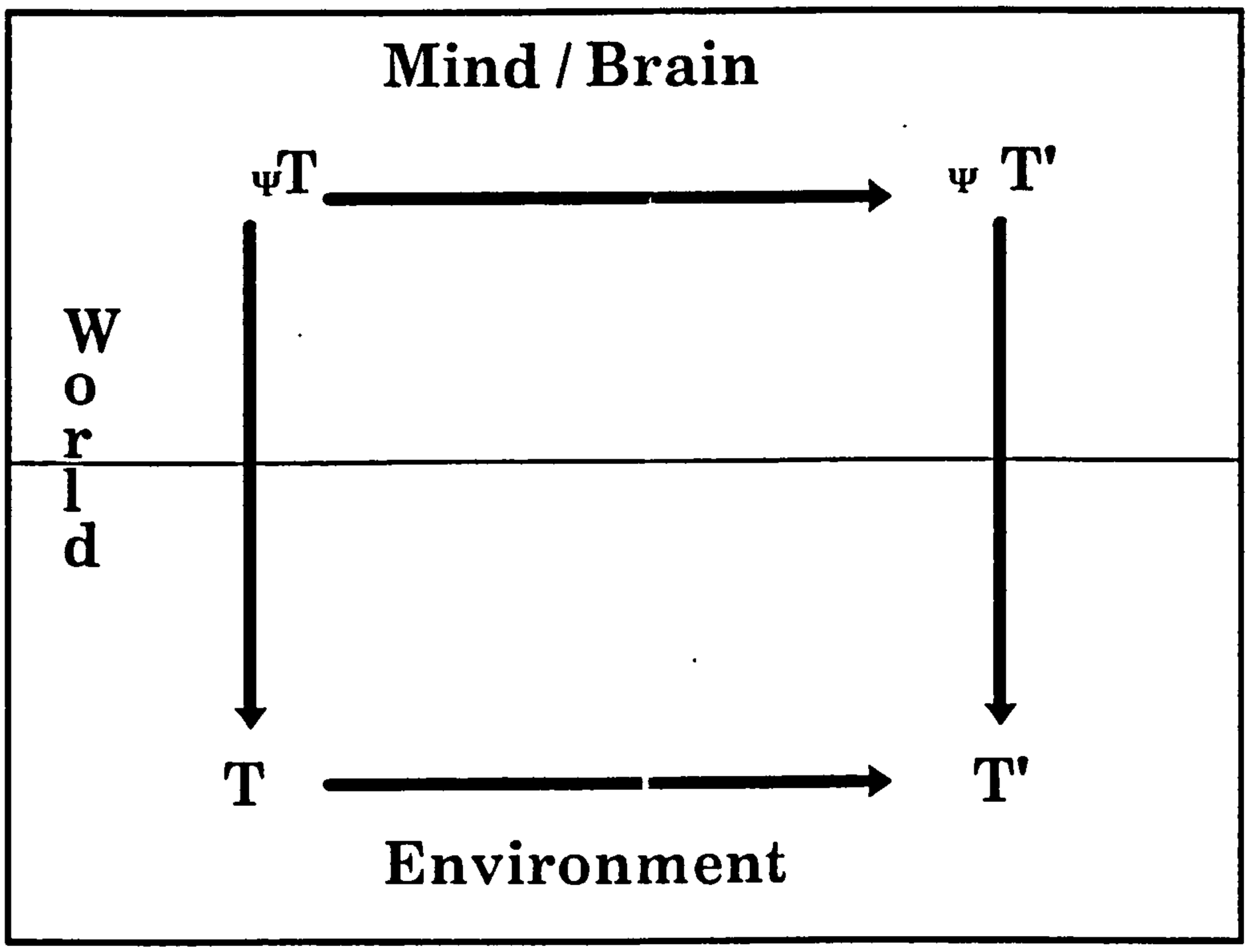

Diagram 2.1 Inference and Information Gain in Situation Semantics.

corresponding constraint ΨC′: ΨT => ΨT′ being actual. This constitutes *a*'s disposition to infer T′ on learning T and is what keeps "his mental state 'in synch' with reality" (Barwise, 1984:23).

a will *gain* information if the diagram commutes ie. on transiting to mental state Ψσ′ of type ΨT′ from Ψσ of type ΨT, T′ is actual. This will depend on whether the background types are appropriately anchored in the situation. If T′ is not actual then, this carries the information that one of the background types did not hold. This information could initiate *inquiry* into which background type it was. The order of search may be determined by familiarity with possible reasons for a constraint not holding. This represents moving into the domain of psychology and process. Diagram 1 can be used to characterise the central problem for a psychology which views transitions between mental states to be a computational process. Distinct from Barwise's (1984) treatment, a clear separation is best retained between the transitions which occur between mental states and constraints. The cognitive system must be able to make various transitions from one mental state to another.

However, for these transitions to constitute effective strategies for acting in the world, they must commute with the operative constraints. It is misleading to characterise the transitions as constraints themselves. The nature of the mechanisms which support inference, ie. the transitions between mental states which systematically track the operative constraints in the world, is a matter for psychology.

Real world constraints purportedly specify a dynamics which would, in the right circumstances, involve transitions from soas of type T to soas of type T′. This is because they specify something about the actual dynamic structure of the world. However, the fundamental problem for a computational psychology is to specify a similar dynamics for transitions between mental states. Two questions are raised. First, what are the mental mechanisms which allow mental states to track real states and the transitions between them in the world? Second, what is it to become "attuned" to the operative constraints in the world such that mental states can be seen to respect the semantics being developed in situation theory. The traditional answer to the first question is mechanised proof theory (Fodor, 1975; Fodor, 1987, Fodor & Pylyshyn, 1988). In the conclusions the viability of this approach will be questioned. The Fodorian position also adopts an answer to the second question (Fodor, 1975, 1980) which will be discussed in the next chapter.

The traditional solution is also closely allied to a position which radically questions the coherence of the view outlined in diagram 1. Empiricism holds that it is fundamentally misguided. The only real transitions which occur are the psychological inferential ones. "Real world" transitions are merely *projections* onto an unstructured world of our tendencies to make inferences. As a first move in the direction of answering some of the psychological questions posed above, the tension between these positions needs to be resolved. Their resolution will fundamentally affect the way in which the psychological questions about learning and inference are addressed.

Summary

In subsequent chapters the formalism of the theory will prove less important than the conceptual foundations which will be discussed in the next chapter. The technical detail will be required however. In, for example, chapter 5 some detailed worked examples will be employed to render absolutely explicit how the situation theoretic view of inference as information gain permits a rational foundation for psychological findings on human conditional reasoning. These accounts will also rely upon the distinction between taxonomic and

non-taxonomic constraints. Their principle role will be to encode the psychologically important distinction between thinking about instances or single occurrences of events possessing two properties and discrete events being related by higher order relations. The manner by which argument roles permit the encoding of the relevant contrast classes defined by a negative soa will also prove central to providing a rational basis for psychological behaviour on conditional reasoning tasks when negations are systematically permuted between antecedent and consequent. A similar role will be played by the preclusion relation.

In the next chapter the concept of a constraint is located in the philosophical literature. This has two purposes. First, to justify the concept. Second, to show its relevance to the issues surrounding scientific hypotheses, but more particularly scientific laws and their confirmation, which are of central concern to the principle psychological task which will be investigated in subsequent chapters.

Chapter 3: The Theory Justified

3.1 Introduction

Problems for situation semantics derive from the fact that the model outlined at the end of the last chapter flies in the face of the last 300 years of philosophy. Due to the empiricist theory of meaning adopted by Hume (taken from Locke) a deep seated skepticism was induced concerning the existence of objective causal dependencies in the world. Unlike individuals, relations were not ontologically respectable. They could not be treated as primitive but were to be analysed or otherwise reduced to set theoretic constructs, to probabilities, or to psychological somethings. A second problem, relates to the move made by situation semantics to treating the interpretation of all conditionals as proposing the existence of "constraints", ie. law-like, explanatory or causal relations. While it may be reasonable to treat *some* conditionals as asserting the existence of relations, eg. those imputing a causal connection between antecedent and consequent, it seems a considerable leap to treat *all* conditionals as referring to *real world* relations.

Two questions need to be initially separated. First, can the skeptic be answered over the need for objective causal dependencies? Second, can Barwise and Perry legitimately bootstrap up to the position whereby all conditionals can be treated as referring to dependencies of a common species with causality? These questions will be dealt with separately and in the order indicated. It will be found that being able to answer the second question in the affirmative relates rather directly to having something to say about the role of *experience.* The relations of which people have experience legitimise their inferences. This involves the old Humean problem of *Induction.* The question posed is what role does induction play in learning and can this process be used to justify our use of knowledge of these relations to make predictions? It will be discovered that providing an alternative answer to Nelson Goodman's New Riddle of Induction provides the answer to whether those boots can be appropriately strapped.

Since Hume, many attempts have been made to reduce statements about causal relations to some other less problematic concept. The arguments that will finally establish the need for objective causal dependencies rely on showing that two philosophically "received" views of how to achieve this reduction entail a circular appeal to the concepts of causal laws,

dependencies or powers. However, Hume is the starting place. What he had to say not only established the problem space, but will also figure in part of the resolution of the problem. Care will have to be taken in assessing arguments as philosophical standards change over the centuries and have indeed changed recently. The issue concerns the demand for *analysis* as opposed to possibly *psychologistic* responses to philosophical problems. This issue will be discussed as and when it arises.

The problem is to make sense of what is meant by the use of particular causal statements, let e_1, e_2 be token events:

(3.1) e_1 caused e_2

Attempted solutions involve analysing many other related statement forms and concepts. (i) Universally quantified conditionals, (ii) counterfactual conditionals, (iii) scientific laws, (iv) law-like relations, (v) dispositions all form a conceptually related group. All have been appealed to in attempting a reduction of (3.1). (3.1) is either subsumed under the generalities in (i), (iii) or (iv), or is synonymous with (ii) or (v). However, I will begin with Hume's psychological reduction. I will then look at Goodman's (1983, originally, 1955) reformulation of the problem and his attempted reductions via (ii) and (iii), and the emergence of his new riddle of induction. I will then pause to consider, Goodman's and Fodor's responses to the riddle, where some discussion of the "demand for analysis" will be required. I will then outline two recent positions, one due to Cartwright (1983) and one due to Stalnaker (1984), which argue for objective dependencies in the manner described above. I will then go on to discuss the problems created for situation semantics due to the boot strapping maneuver. This will be followed by an attempt at a resolution of these problems.

3.2 The problem of objective causal dependencies

3.2.1 Hume

The problem Hume identified was that of justifying the grounds upon which people take the causal relation to warrant inferences concerning future events. ie. predictions (cf. Ayer, 1980). If the grounds are purely inductive, ie. I have observed that every B-occurrence has been preceded by an A-occurrence, then I have no grounds to infer that on the next A-occurrence a B-occurrence will follow. Any appeal to the future conforming to the past etc. is doomed to circularity by the need to appeal to the very inductive grounds such a premise is supposed to justify. However, it seems that people are led to predict B-

occurrences on observing A-occurrences by some *necessary connection* between A-occurrences and B-occurrences. This can not be logical necessity as it does not imply a logical contradiction to deny that a B-occurrence will follow an A-occurrence on the assumption that they are causally related (cf. Bennett, 1971:272). Moreover, logical necessity holds only between propositions not to the facts therein described. Because of his empiricism, Hume was lead to argue that the concept of necessary connection in the world (ie. causal necessity) was unintelligible.

The theory of meaning Hume adopted was essentially Locke's. The meaning of a term was simply the "idea" it was associated with for an agent. For Hume, ideas were simply weaker, less vivid, versions of "impressions" which were the immediate data given by the senses. He also adopted the simple/complex distinction introduced by Locke. A term could either have a complex meaning or a simple meaning. If a terms meaning was complex, then it was decomposable into simples. If a terms meaning was simple it was associated directly with an idea, which was a less vivid version of an immediate sense impression. All terms are meaningful only to the extent that they are either simple or complex ideas. For Hume, the concept of "necessary connection" was unintelligible because (i) it could not be assigned a complex meaning, ie. it could not be decomposed into simpler parts; and (ii) it could not be assigned a simple meaning, ie. there is nothing in the impression of, for example, one billiard ball making another billiard ball move which corresponds to the necessary connection. To make sense of particular causal statements Hume, therefore, had to proffer a reduction of e_1 caused e_2 which accorded with his semantic theory.

All that was contained in the sense impression was the relata, and the impressions of *contiguity*, and *priority*, ie. the relata are "constantly conjoined". The question Hume addressed was what needed to be added in order to account for the feeling of necessitation we subjectively experience in making predictions? Two factors conditioned Hume's reduction: (i) his aforementioned Lockean, ideational, or empiricist theory of meaning, and (ii) his "geneticism". A genetic approach can be contrasted with an analytic approach. On the former, an account of, say, "cause" will involve outlining what it is that leads or causes us to make predictions. On an analytic account one is required to outline in what a concept, such as cause, consists, without implying anything about what leads people to predict (the latter is relegated to the level of psychological explanation). Hume finds the third ingredient in the self evidence of the following claim which establishes that people can have no more than inductive grounds for making predictions:

(3.2) "... even after the observation of the frequent or constant conjunction of objects, we have no reason to draw any inference concerning any object beyond those of

which we have had experience" (Hume, *A Treatise of Human Nature*, p. 139).

Hume's solution "is that the observation of the frequent or constant conjunction of facts of recurring types gives rise to a habit or custom of expecting this regularity to be repeated" (Ayer, 1980:66). The repeated observation of conjoined events habituates people to expect that a regularity will be repeated. The habit or custom, becomes so entrenched that people *project* the subjective association of ideas onto the world. Hence the belief arises that "necessary connections" hold between the relata themselves. This is the first statement of the empiricist *Projection* strategy regarding relational structure in the world. Causality and like relations are reduced to projections of the structure of the mind onto an unstructured reality.

There are many points of detail which can be questioned in Hume's response. For example, his ideational semantic theory which drive him inwards to look for the third ingredient rather than outwards to natural laws perhaps. His assumption of the passivity with which people arrive at habits of inference which may equally be arrived at by active and conscious cogitation and conscious acceptance of a custom. The very notions of contiguity and priority no longer seem essential to the concept of "cause". No modern philosopher would wish to beg the question against backwards causation, nor contemporary scientist give up the notion of action at a distance (cf. Bennett, 1971). However, the reductive projection strategy has survived the years since Hume and this is in no small measure due to the synthesis of empiricism and rationalism bought about by Kant.

3.2.1.1 A Kantian digression

Hume's position is often characterised as a skeptical solution to a skeptical problem, ie. he proffers a psychological reduction of a philosophical problem (Kripke, 1982). In effect he gives reasons why people confidently predict on the basis of inductive evidence but no justification of this procedure without which it could not guarantee certain knowledge. The solution proposed by Kant runs as follows.

Kant was one of the first philosophers to realise that there is an active component to perception. Conscious experience of sensations involves the application of *a priori* intuitions concerning, eg. space and time, which actively structure experience. These intuitions are required to allow conscious appreciation of even *subjective* sensations which Kant called "representations". In virtue of these intuitions, however, only representations would be possible. *Objective* experience of an external world was only possible given that a persons

representations were further structured in accordance with certain *a priori* concepts, eg. the concept of necessary causal connection:

(3.3) "So, Kant held that we could only know the empirical world as something interpreted in accordance with "rules" which were innate in our minds". (Trusted, 1979:56)

Shorn of the distinction between *noumena* and *phenomena* and the goal of certain knowledge, the picture of the mind structuring its representations of reality in accordance with innate rules is one with which a modern cognitive scientist would be wholly familiar. Indeed Fodor (1985), has advanced the opinion that the only significant advance made by cognitive science over the views of Kant and the British Empiricists has been the computer metaphor. Fodor, like Kant, subsequently proffers a rationalist solution to an empiricist's skeptical problem in the *Language of Thought* (1975).

3.2.2 The "Humean" view

Jonathan Bennett (1971) observes that the modern "Humean View" of the causal relation arises from making Hume's psychological reduction analytically "respectable". This is achieved by removing the genetic, psychological component from the theory by appealing to the genetically neutral concept of a "disposition". Where "disposition", "is to be understood in Ryle's [1949] way: to credit someone with a disposition is to speak not of what he feels like doing but only of what in *general* [my italics] he does do or would do if..." (Bennett, 1971:305). This is in part a contemporary empiricist retraction of the Kantian picture based on the logical behaviourism of Ryle. However, the problems that emerged at around the same time for the Humean view led to Fodor's (1975) re-instantiation of this picture. Now made analytically respectable, the modern Humean view has the following form:

(3.4) "The difference between 'e_1 caused e_2' and 'e_1 preceded e_2' is that the former entails that there is a law which..." (Bennett, 1971:307)

In this context a "law" is a true, contingent, universally quantified conditional statement. This recognises the important element of generality observed in the move to dispositional terminology, ie. particular causal statements are subsumed under, or are particular instantiations of general laws.

This analytic response is intentionally divorced from psychological and epistemological concerns. It has nothing to say about how humans represent and utilise lawlike relations,

nor about how people come to know that any particular contingent, universally quantified conditional is true. "Laws" are simply identified with universally quantified conditionals; ie. statements of the form ∀x(Fx -> Gx). This is a reduction of particular causal statements to instantiations of general laws of association, ie. things of type F are generally associated with things of type G. No relation between things of type F and things of type G, apart from inclusion between classes, forms part of the analysis, thereby effecting the desired reduction.

3.2.3 Goodman

The observation that the e_1 caused e_2 locution possesses a subjunctive character lead to Goodman's reassessment of the analytic reduction, ie. because of this additional twist the reduction could not be direct. Making the Humean View of causation "analytically respectable" involved introducing the concept of a disposition. This concept was introduced in Bennett's adversion to Ryle:

> (3.5) "...this is to be understood in Ryle's way: to credit someone with a disposition is to speak not of what he feels like doing but only of what in general he does do or *would do if* [my italics]..." (Bennett, 1971:305).

Science, in general, also makes appeal to dispositional terms like "solubility", "fragility" etc. These terms possess a subjunctive character, in that they seem to describe "not only what *has* happened or what *will*, but also describe what *would* happen under various circumstances" (Suppe, 1977:36). This subjunctive character is also a factor in Ryle's psychological application of the term, ie. dispositions refer to "what in general he does do *or would do if...*". Particular causal statements also have this subjunctive character. Hume's famous "second" definition of causality was framed in terms of a *counterfactual* conditional (I adopt the usual philosophical convention of using "counterfactual conditional" to refer to all conditionals in the subjunctive mood, regardless of whether the antecedent is contrary to fact):

> (3.6) "...we may define a cause to be an object followed by another, and where all the objects, similar to the first, are followed by objects similar to the second [first constant conjunction definition].
> Or, in other words, *where, if the first had not been, the second never had existed* [second counterfactual definition]" (Hume, *An Inquiry Concerning Human Understanding*, p. 87)

Treating e_1 caused e_2 and

(3.7) If e_1 had not occurred, then neither would e_2.

as synonyms, the question becomes can a semantics, ie. truth conditions, be provided for (3.7) without appeal to causal locutions. Chisolm (1946) had already observed that dispositions could not be provided with truth conditions using just the material conditional (ie. using reduction sentences). However, at the time it was a desiderata to provide a purely extentional analysis. If nothing else this was because the logical positivists received view (the earliest statement of which is, Carnap, 1923; for later formulations, cf. Hempel (1952) and cf. Suppe (1977) for an overview) held that scientific theories were axiomatisable using just the first order calculus. The problem that arises for dispositions is as follows. "Fragility", for example, could be defined as in (3.8):

(3.8) X is fragile iff, if X were dropped, then X would break

The problem for an extentional account of this conditional, ie. using the semantics for the material conditional, is that the right hand side of (3.8) would be true of any X which was not dropped. However,

(3.9) If X were dropped, then X would *not* break

would also then be true of any such X. So, it would appear that counterfactual conditionals cannot be analysed using only the material conditional. Goodman felt that he might circumvent this problem by invoking the concept of law like relation still analysed as above and thereby provide an extentional analysis of counterfactuals.

Goodman notes that the truth of counterfactuals seems to require that a certain connection obtains between antecedent and consequent. However, the consequent does not follow from the antecedent by logic alone. Take the following example:

(3.10) If the match had been scratched, it would have lighted

Asserting (3.10) amounts to the claim that the consequent can be inferred from the antecedent if the relevant conditions, eg. the match is well made, there is sufficient oxygen, the match is dry etc., hold. However, the conjunction of all the relevant conditions with the explicitly stated antecedent does not yield the conclusion that the "match lights" as a logical consequence. It only follows in virtue of some natural law holding between these conditions and matches lighting.

As the truth of a counterfactual seems to depend on the elliptically present relevant conditions in the antecedent, Goodman first attempts to specify truth conditions for

counterfactuals by placing restrictions on the conjunction of the relevant conditions with the antecedent. So the general form of the counterfactual is take to resemble the implicit form of a default rule:

(3.11) p & RC -> q

Goodman's classic argument is that one of the conditions which must be satisfied is that p must be *cotenable* with RC; where p is cotenable with RC if it is not the case that RC would not be true if p were. However, this leads to an infinite regress: to determine whether p and RC are cotenable involves determining whether the counterfactual "if p were true then RC would not be true" is itself true. Cotenability is itself defined in terms of counterfactuals, so the truth of a counterfactual always involves determining the truth of another counterfactual ad infinitum. So, an analysis of counterfactuals will have to be in terms of the natural laws which license the inference from p & RC to q. (The introduction of relevant conditions serves to indicate the quite radical context sensitivity of the counterfactual and/or the laws which support them.)

This leads to the second half of Goodman's project: the analysis of law-like relations. Goodman adopts the view stated above that such a relation is analysed as a universally quantified material conditional. However, not all generalisations are law-like. For example:

(3.12) All ravens are black.

(3.13) All the coins in my pocket today are silver.

Two factors distinguish (3.12) from (3.13):

(i) To confirm (3.12) one would not need to examine all ravens, in Goodman's terminology (3.12), but not (3.13), is a *projectible* hypothesis.

(ii) Due to (i), (3.12) (but not (3.13)) could be used to predict the colour of as yet unobserved ravens (coins).

Both (3.12) and (3.13) are based on observed empirical regularities and so on the Humean View both should give rise to habits of inference. But only (3.12) can be reasonably expected to do so. So, *not all* empirical regularities give rise to habits of inference, hence the *New Riddle of Induction* is how to distinguish between those that do from those that do not.

Goodman considers the old Humean problem of the justification of induction to be solved. People are as justified in using inductive procedures as they are deductive procedures. The

only justification each has is that they conform to peoples' respective inductive and deductive practices. The only issue which separates them is that deduction is formalised in a calculus which guarantees confidence in deductive practices. The new problem of induction is to specify a similar calculus which allows a principled distinction between law-like (3.12) and accidental generalisations (3.13), projectible hypotheses from non-projectible hypotheses. Such a calculus is the purview of *Confirmation Theory*. So, the reduction of particular causal statements now relies on whether sense can be made of confirming universal laws of association while retaining a principled distinction between law-like and accidental generalisations. The pure analytic reduction is now making appeal to epistemology. But this is analytically respectable if a *formal* account of what it is for a universal law to be confirmed can be provided.

Devising a logic involves formally defining a *consequence* relation. This can be done syntactically or semantically. Perhaps by analogy it would be possible to formally define a *confirmation* relation. A straightforward proposal would be to treat confirmation as the converse of deduction, ie. as deduction in reverse The desired relation of confirmation can be defined as follows:

(Conf.) If A |- B and B, then C[B,A] (read, "B confirms A")

ie. a hypothesis A, is confirmed by the truth of its deductive consequences B (Hempel, 1965; Carnap, 1950). This suggestion, however, leads to the many paradoxes of confirmation theory. If the plausible condition is added that any evidence *e*, which confirms a hypothesis *H*, also confirms the logical consequences of *H*, ie. the special consequence condition:

(Cons.) (C[*e*,*H*] & *H* |- *H'*) -> C[*e*,*H'*]

then the conclusion can be derived that every hypothesis confirms every other hypothesis. The proof is trivial: take any hypothesis H_1; this is a consequence of the conjunction of H_1 with any other hypothesis, say H_2, and thereby confirms this conjunction by the present criterion. This confirmed conjunction, H_1 & H_2, has H_2 as a consequence, therefore any hypothesis, H_1, confirms any other hypothesis, H_2.

This conclusion can be avoided. It may be the case that although all statements which confirm a hypothesis are deductive consequences of it, it is not the case that all the consequences of a hypothesis confirm it. For example, although finding out that John is Greek supports the hypothesis that John is a tall, overweight, Greek shipping magnate, "by reducing the net undetermined claim" (Goodman, 1983:69), it is counter-intuitive to argue that

this observation confirms this particular hypothesis. There is no transfer of credibility either to the other components of the claim or to other instances. This indicates that the claim should be relativised to instances mentioned in the general claim. This is equivalent to restricting the domain of the universal quantifier in a hypothesis to the kinds of thing mentioned in the antecedent. Restricting the domain of the quantifier in this way is called the "Sufficiency Condition" or "Nicod's Criterion". Thus, ∀x(Fx -> Gx), is interpreted as: for all things x of kind F, Gx. (This form of restriction on the quantifier is a *property* restriction, ie. a restriction to still potentially infinite set of a certain kind of object. Later on the possibility of *objectual* restrictions will emerge, ie. restrictions to finite domains, cf. chapter 4)

Take the hypothesis that ψ: *All ravens are black* and render it logically in the familiar way: ∀x(Ax -> Ax). Then take an instance *a*, by restricting the domain of the quantifier: Fa & Fa confirms, Fa & ¬Fa disconfirms, and, ¬Fa & Fa and ¬Fa & ¬Fa are irrelevant. Nicod's criterion seems admirably sensible. But a problem is created if another reasonable adequacy condition on the relation of confirmation is assumed, ie. if *e* confirms H_1, and H_1 is equivalent to H_2, then *e* confirms H_2, ie.

(Equiv.) C[*e*,H_1] & (H_1 <-> H_2)) -> C[*e*,H_2]).

This is called the *equivalence condition* and it leads directly to the notorious *Ravens Paradox*:

(Ravens) ∀x(Ax -> Ax) <-> ∀x(¬Bx -> ¬Rx), and
C[¬Ra & ¬Ba, ∀x(¬Bx -> ¬Rx)], therefore, given (Equiv.)
C[¬Ra & ¬Ba, ∀x(Rx -> Bx)]

That is, instances of non-black non-ravens, eg. white plimsoles, confirm the hypothesis that all ravens are black, or more generally any evidence which does not disconfirm a hypothesis confirms it. So, even if Nicod's criterion is accepted subject's should turn *all* the cards if they were confirming (confirmation will now be contrasted with verification) as ¬*p*, ¬*q* instances also confirm.

(Most attempts to get round this problem involve rejecting the equivalence condition on the relation of confirmation. However, another way out is to reject the material conditional. Perhaps it is the particular equivalences licensed by this connective which cause the problem. This route has been explored by Belnap (1970) who defines a connective he calls *conditional assertion* ("/"). This is based on the Quine-Rhinelander idea that a conditional is not an assertion of a conditional but a conditional assertion, ie. the consequent is affirmed

only on the assumption that the antecedent is true. On the semantics Belnap provides "/" does not contrapose. Therefore, the equivalence which the ravens paradox relies upon does not hold, and hence the paradox does not arise; cf. chapter 4, on "Defective truth tables").

The paradoxes which beset confirmation theory lead Goodman to reject any formal attempt to derive a confirmation relation. He further drives home his skeptical point by demonstrating that any hypothesis which was first felt to be projectible on the basis of the evidence to date, can be turned into one which is obviously not so projectible. By locating the problem in the nature of the predicates used to categorise the world Goodman simultaneously rejects any simple formal solution to his skeptical puzzle. Take the following hypothesis:

(3.14) All emeralds are green.

This looks like a perfectly respectable projectible hypothesis. However, take the predicate "grue", such that:

(3.15) $\forall x$(x is grue iff x is green before 2000 a.d. or x is blue on or after 2000 a.d.)

then the hypothesis that:

(3.16) All emeralds are grue.

is as equally confirmed by the evidence to date as (3.14), but (3.16) is obviously not projectible. After 2000 a.d. a green emerald would no longer be a grue emerald. This poses the problem of how one can know whether "green" does not function like "grue"? Observing, for example, that "green" is not disjunctive fails to resolve the problem because whether a predicate is disjunctive or not is relative to the predicative base one starts with. On a predicative base containing grue-type predicates, "green" would be a disjunctive predicate. There have been many attempts to circumvent these problems. However, Hilary Putnam notes in his forward to the fourth edition of Goodman's *Fact, Fiction and Forecast* (1983) that none appear to succeed in their goal of permitting a formal definition of inductive validity. Induction relies on content. ie. the concepts/predicates by which people categorise their world.

3.2.4 Two "solutions"

Two "solutions" to the problem are prominent in the literature: one due to Goodman himself, the other due to Fodor (1975, 1980). Both could be described as skeptical solutions to skeptical problems in that they invoke non-analytic psychological considerations about what

causes or leads to a hypothesis being projectible. Fodor's response mirrors Kant's to Hume. Direct appeal is made to innate concepts or predicates and an innate ordering of hypotheses which structure our experience of the world. Insofar as Kant's original answer to Hume was convincing as a rationalist justification of induction one may be persuaded that Fodor's response offers a similar justification in the face of Goodman's riddle. Fodor's position, however, is less of a justification of induction than a denial that induction plays any significant role in learning. Hypotheses can not be induced from experience because the *a priori* possession of the appropriate concepts (predicates) is a pre-requisite for experience. Learning is the process of confirming or disconfirming hypotheses which are given *a priori*, ie. innately. This imbues Fodor's response with an analytic flavour. It apparently makes no appeal to particular matters of fact, but relies on arguments concerning what people require to have any experience of particular matters of fact at all. This relies on the identification of the *a priori* with the innate. However, what is innate and what is not is an empirical question, *not* to be determined *a priori*. And so Fodor's response is on the same psychological footing as Goodman's.

Goodman proffers a typically pragmatic response. Conceding the circularity, he allows that the predicates which are now considered projectible are those which have been previously projected within a linguistic community. These predicates have become "entrenched" in that community and although guaranteeing nothing about future projections they none-the-less conform with current practices. And that is all that can be legitimately asked for.

One consequence of pursuing an analytic reduction, which captures the subjunctive character of causal statements, is that in the limit genetic, psychological or *intentional* factors have been invoked. The empiricist *projection* strategy (not to be confused with the "projectibility" of predicates) is retained in both responses. For Fodor, objective dependencies are simply the projections onto the world of peoples' well confirmed innate hypotheses. For Goodman, they are the projections onto the world of peoples' projectible habits of inference, where these are given by the well entrenched predicates of a persons linguistic community.

3.2.5 The demand for analysis

Both responses appear to violate the demand for analysis. That is, the demand to resolve an epistemological problem using the tools of metaphysical analysis. One is allowed the tools to specify in what a concept must consist without making appeal to how the concept may

actually arise for an agent with thus and so cognitive resources. Denying the legitimacy of this demand forms part of Fodor's (1975) criticism of Ryle (1949). Separating conceptual or analytic responses from causal or genetic ones, Fodor views as a mistake. Dennett's (1978) criticism of Fodor amounts to an accusation of committing the fallacy of *ignoratio elenchi*, ie. the proposition Fodor refutes (the postulation of inner cognitive processes plays no explanatory role in psychology) is not the proposition defended (the postulation of inner cognitive processes plays no role in philosophical analysis). Ryle is making a philosophical point, not a psychological one. As Dennett observes, Ryle's analyses of cognitive concepts are replete with intentional idioms. Viewed from a historical perspective the apparent problem can be resolved. The view that philosophical analysis was independent of science, and of psychology in particular, dominated when Ryle was writing. Fodor is observing that this is not the case. And this is the contemporary view, especially in epistemology (Goldman, 1986) but also in metaphysical analysis. For example, Stalnaker (1984) objects to Lewis' (1973) conception of the methodology of metaphysics as succeeding:

(3.17) "...to the extent that (1) it is systematic, and (2) it respects those of our pre-philosophical opinions to which we are firmly attached."

insofar as it should:

(3.18) "...also help to explain the source of those opinions and their role in our practical activities" (Stalnaker, 1984:50).

Within the context of peoples' pre-philosophical opinions concerning modality (insofar as they relate to an understanding of objective dependencies in the world), Stalnaker interprets (3.18) as the demand to show how a possible worlds semantics for counterfactuals relates to some conception of an intentional state. He attempts to show how an abstract semantic analysis of the counterfactual can inform conceptions of what constitutes a law-like relation (Stalnaker, 1984). In doing so, he makes explicit appeal to the empiricist projection strategy.

By more contemporary lights, either of the proposed responses are respectable. However, Goodman's "solution" mirrors much of the later Wittgenstein. The properties, eg. meanings, which attach to words are derived from their place within a communities "language games" or customs and practices which are described in language. One language game involves predicting future events, and the predicates chosen in this "game" are those which will be projected. This view of language is not the view adopted by Cognitive Science which has its roots firmly in the early Wittgenstein and the heirs of the *Tractatus*, eg. Carnap, Tarski, Davidson etc. To the extent that this is so, the Fodorian response is the "only straw afloat",

to quote an oft used phrase. However, Fodor's innatist response is just one other empirical hypothesis about the nature of the device. The best response to an *a priori* claim to innateness is to come up with an alternative straw to grasp concerning the mechanisms make up. As a corollary to exploring the empirical adequacy of an alternative semantic position (situation semantics), some speculations on mechanism will be proffered in the conclusions.

Despite the contemporary respectability of these responses, there may be no need to go as far as either Goodman or Fodor. Both result in a psychological reduction due to the attempt to resolve new problems arising with induction. However, perhaps it would be conceptually more satisfactory if the semantic question of the meaning of particular causal statements and the epistemological problem of induction could be kept separate. To do so would involve stopping the reduction at some earlier point. This has been attempted in two ways. First, the observation that most laws of association are probabilistic may allow a reduction by placing sufficient restrictions on the probabilistic relation (conditional probability) which putatively holds between e_1 and e_2. Second, the contemporary gloss on law-like generalisations as those which support counterfactuals, could be captured by providing an intensional semantics for the counterfactual. It emerges that either way the reduction is attempted involves circular appeal to causal laws, powers, or properties. Although this may indicate that Goodman's and Fodor's responses still stand an alternative will be proposed in section 3.4.

3.2.6 Cartwright

Successful prediction relies on the basis of an organisms predictions constituting *effective strategies* for dealing with its environment, ie. achieving its goals. The concept of what constitutes an effective strategy is an interesting turn in Cartwright's (1983) argument that causal laws can not be reduced to probabilistic laws of association. In effecting the reduction of particular causal statements the proposal is that they are instantiations of general causal laws and the latter can be reduced to probabilistic laws of association. (That is we no longer accept the $\forall x(Fx \rightarrow Gx)$ general form for laws of association). Cartwright's argument for the objectivity of causal laws is somewhat technical but can be summarised quite succinctly.

The argument depends on a quantity known to statisticians as partial conditional probability: Prob($E/C.K_j$), ie. the probability of E given C holding K_j fixed. The main line of the argument is that in using this quantity in attempting to reduce causal laws to probabilistic

laws of association or in defining an effective strategy the K_j that need to be held fixed are all and only the *causally* relevant factors. If the factors that are to be held fixed are determined in any other way there can be no guarantee that the appropriate probability measure (ie. increase in conditional probability) taken to define whether E is a cause of C, or C is an effective strategy to achieve goal G, will be observed.

Some examples will clarify the force of the argument. The following example makes explicit the distinction between an effective and an ineffective strategy: Cartwright received the following letter from an insurance company:

(3.19) "It simply wouldn't be true to say,

> "Nancy L. D. Cartwright...if you own a TIAA life insurance policy you'll live longer"
>
> But it is a fact, nonetheless, that persons insured by TIAA do enjoy longer lifetimes, on average, than persons insured by commercial insurance companies that serve the general public." (Cartwright, 1983:22)

Agreeing with the sales pitch, buying a TIAA policy would not be an effective strategy for lengthening ones life, but perhaps stopping smoking would be. If this is so, then the difference must depend on "on the causal laws of our universe, and on nothing weaker" (Cartwright, 1983:22). This can be seen in two ways (i) in the attempt to provide a statistical analysis of causation, (ii) in attempting to provide a statistical analysis of an effective strategy.

(i) The intuition behind a statistical analysis of causation is that if, say, generally smoking (*S*) causes heart disease (*H*) (note '*S*' and '*H*' = event *types*), then the conditional probability, Prob(*H*/*S*), is greater than the probability of *H* alone, Prob(*H*), ie. Prob(*H*/*S*) > Prob(*H*). However, if smoking is also correlated with a sufficiently strong preventative factor, say exercise (*X*), then the expected increase in probability will not appear. As long as in the population smoking and exercising are sufficiently highly correlated, then any change in Prob(*H*/*S*) can be counterbalanced by the preventative effects of factor *X*. Generally, all counterexamples to the claim that causes increase the probability of their effects involve showing that there is some other *causal factor* which dominates in this particular case. So, the only way out is that the population should be selected such that, with the exception of *C*, it is causally homogeneous with respect to the effect. Cartwright uses Carnap's concept of a *state description* to pick out the causally homogeneous population, a technical twist I will not elucidate further. The definition is as follows:

(3.20) C caused E iff Prob($E/C.K_j$) > Prob(E/K_j).

Where K_j is the set of state descriptions over $\{C_i\}$, and $\{C_i\}$ satisfies certain requirements:

(a) If $C_i \in \{C_i\}$, then C_i causes or precludes (prevents) E.
(b) C is not included in $\{C_i\}$.
(c) $\forall D$(if D causes or precludes E, then $D = C$, or $D \in \{C_i\}$).
(d) If $C_i \in \{C_i\}$, then it is not the case that C causes C_i.

Apart from (d), these are self explanatory, (a) guarantees causal relevance, (b) ensures that only the relevant factor C, is heterogeneous in the population, and (c) ensures *all* causally relevant factors are included. (d) is there to ensure that any factors in the causal chain from C to E are not held fixed. This is of course circular as a reduction of C causes E, as this locution occurs on both sides of the definition. All other *causal* factors have to be identified in order to establish whether one factor possesses the desired statistical property.

(ii) A directly analogous situation emerges in defining an effective strategy for reaching a particular goal in decision theory. Again the idea is that if S is an effective strategy to reach goal G, the conditional probability, Prob(G/S) should be greater than the probability of G alone, Prob(G), ie.

(3.21) S is an *effective strategy* for G iff Prob(G/S) > Prob(G.

The problem again arises that in "populations where the strategy state is correlated with other factors causally relevant to the goal state" (Cartwright, 1983:34) the conditional probability will fail as a good measure of effectiveness. So again it must be ensured that the population is not causally heterogeneous with respect to the goal state, apart from the particular strategy in question. So, identical restrictions have to be placed on (3.21):

(3.21') S is an *effective strategy* for obtaining G in situation L iff Prob($G/S.K_L$) > Prob(G/K_L).

All the above restrictions (a) - (d) then apply to $\{C_i\}$, ie. the causally relevant factors for G. And again this involves appeal to causal properties and therefore defining an effective strategy makes ineliminable appeal to objective causal dependencies.

The reduction of particular causal statements via general causal laws and probabilistic laws of association fails. The question which Cartwright takes the turn into effective strategies to answer, is what difference do casual laws make? They make a crucial difference to the effectiveness of the goal directed strategies of an agent. Cartwright does not deny that statistical analysis can aid in discovering causes, it is just that to do so presupposes causes are already objective features of the world. Perhaps, then the counterfactual route can proffer a

reduction? It is interesting, but perhaps not all that surprising, that the argument that leads Stalnaker (1984) to the view that causal dependencies are objective features of the world sounds very similar to Cartwright's.

3.2.7 Stalnaker

Stalnaker (1968) defines a new conditional connective (">") in which the content of the antecedent and consequent remain the same as for the indicative but the subjunctive mood of the verb is moved into the conditional:

(3.22) If a match is struck it lights. (indicative mood, unmarked):
(a match is struck) => (it lights)

(3.23) If the match were struck, it would light. (subjunctive mood):
(a match is struck) > (it lights)

"=>" is used to distinguish the indicative from the material conditional ("->"), as it is unlikely that the latter is a sound analysis of the former (Nute, 1984). Stalnaker then defines a possible worlds semantics for this connective, ie. they are treated as dyadic modal operators. The models employed are extensions of the monadic case. Getting the formalities over with quickly will permit a ready understanding of Stalnaker's argument.

A simplified Stalnaker model (Nute, 1984:397) is an ordered quadruple $\langle I, R, s, [] \rangle$, where I is a non empty set of possible worlds; R is the binary accessibility relation which is reflexive in Stalnaker's model; s is a world selection function which assigns to a subset A of I and a member i of I a subset $s(A, i)$ of I; [] is a function which assigns to each sentence ϕ a subset $[\phi]$ of I, $[\phi]$ is then said to be the proposition expressed by ϕ. The truth conditions of a conditional $\phi > \psi$ are defined as follows:

(i) $\phi > \psi$ is true iff $\varnothing \neq s([\phi], i) \subseteq [\psi]$.

(ii) $\phi > \psi$ is false iff $s([\phi], i) \not\subseteq [\psi]$.

(iii) $\phi > \psi$ is undefined iff $s([\phi], i) = \varnothing$

In some accounts (Lewis, 1973) $\phi > \psi$ is treated as true in the case where $s([\phi], i)$ is empty.

The core of this semantics is the grounds on which s selects worlds in which to evaluate a conditional. The idea is that it should select worlds which are as similar to i as possible in

which ϕ is true. Let us take the example above to illustrate what is going on:

(3.24) If the match were struck, it would light.

Informally the idea is that taking *i* to be the real world in which the match is not struck, then (24) is true if in all worlds most similar to *i* except that the match is struck, the match lights. That is, the consequent is evaluated in all and only the worlds most similar to *i* except for the fact that the match is struck. So the selection function takes as arguments, *i*: the real world in which the match was not struck, $[\phi]$: the subset of worlds in which the match is struck and returns the subset of worlds most similar to *i* in which the match is struck, ie. $s([\phi], i) \subseteq [\phi]$. If in all these worlds the match lights, ie. $s([\phi], i) \subseteq [\psi]$, then (24) is true.

Several additional intuitively correct restrictions are place on the items of the model, eg. all the worlds selected by *s* must be accessible from *i*, ie. for all $j \in s([\phi],i)$, $<i, j> \in R$, and if ϕ is true in *i*, then *i* is the only world selected, ie. if $i \in [\phi]$, then $s([\phi], i) = \{i\}$. It can be the case that the range of the selection function is empty, ie. no other worlds are accesible from *i* in which case on Stalnaker's account $\phi < \psi$ is undefined for all ψ at *i*. However, concern centres on the basis on which comparative similarity between worlds can be taken to offer a reduction of particular causal statements.

Two preliminary observations can be made. First, the proposed reduction does not look very promising. The reduction hoped for by Goodman was of a problematic concept, objective causal dependencies, to an unproblematic concept, counterfactuals and scientific laws. However, notions like overall comparative similarity between possible worlds, seems more rather than less problematic than the concept of causal dependencies themselves. Nonetheless, Stalnaker identifies this Humean reductionist project with attempts by Lewis to explicate the respects of comparative similarity that are relevant to assessing counterfactuals. Broadly this is still the old Humean project, ie. the attempt to vindicate empiricist skepticism concerning the idea of necessary connection in the world. Second, in *Inquiry* Stalnaker provides a conceptualist interpretation of his abstract semantic theory which he takes to aid in understanding its relevance to inquiry. Before outlining the argument for objective causal relations, it is important to clarify the main points of Stalnaker's project.

Stalnaker takes *s*, the selection function to be an abstract characterisation of how people transit from one belief state to another. Rather than proof theoretic steps performed on sentences of mentalese, a la Fodor, propositions are to be treated, as in the semantic theory, as subsets of possible worlds. These can be regarded as abstractions from the idea that beliefs

constitute possible ways an agent thinks the world could be. In this respect the selection function constitutes an abstract characterisation of peoples dispositions to alter their beliefs in response to new information. The relation between indicatives and counterfactuals is a central part of the analysis. What Stalnaker refers to as "open" conditional sentences (in the indicative mood) are statements of dispositions ie. methodological policies to alter beliefs. Possessing such a conditional belief involves being disposed to alter ones beliefs in accordance with it. Making explicit what Stalnaker has in mind when he says that belief in conditional propositions involves *projecting* those methodological policies onto the world will involve exampling the main idea.

Suppose I have the following methodological policy:

(3.25) If someone is a Tory, then they are immoral.

then on learning that Nigel is a Tory I will be disposed to believe that Nigel is immoral. Now suppose I believe the conditional proposition that:

(3.26) If Nigel hadn't been a Tory, he would have given more money to the NHS.

is true. There will be all kinds of reasons why I *could* believe this to be true, including of course all my knowledge about the policies of the opposition parties etc. This is indicative of the fact that all my other factual knowledge and other methodological policies will affect how I evaluate a particular counterfactual. However hard it is *not* to project your own (ie. the reader's) beliefs onto the example, only those explicitly stated are operative. I also believe some other things:

(3.27) The act "giving more money to the NHS" is moral.

I also believe that:

(3.28) Only Tories are immoral (subsuming facists etc.), ie. the connective in (3.25) is a biconditional.

Now in evaluating (3.26), I will be disposed on learning the antecedent to select those worlds most similar to the actual world where Nigel is not a Tory. Because of my methodological policies, (3.25) & (3.28), these will be *moral* worlds (the assumption is that *nothing else is operative to determine other respects of similarity and difference*), and because of (3.27) I will conclude that the consequent of (3.26) is true at these worlds. So the value of the function *s* is taken to be those worlds which are consistent with my particular methodological policies. But my actual belief that (3.26) is *true* is due to my *projection* onto the world of (3.25) and (3.28), ie. I take them to be the facts about the (social) world *in virtue*

of which (3.26) *is true.* Whether they are actual facts or not, for me to assert that (3.26) is true at least requires that I believe them to be a factual. (I might add that although I represent these methodological policies as *projections* I think they are facts, which goes to show Stalnaker can't be too far off the mark).

This is precisely where the reductionist can get a foot in the door. There are two possible reductionist moves here: one psychological the other analytic. Stalnaker's conceptualist move means that he could, if he wished, proffer a Humean psychological reduction: all he has done is to be a little more explicit in his characterisation of a habit of inference. Lewis on the other hand, who adopts a realist stance with respect to possible worlds, is offering an analytic reduction. Either way, I only have to *believe* that my methodological policies are factual, which by no means implies that they have a basis in fact. The example was deliberately picked as a blatant piece of prejudice to illustrate the point about projection. Moreover, the example has artificially constrained the respects of similarity that are relevant to *s*. For Lewis' reduction to go through he must show, in general terms, how the relevant respects of similarity and difference can be spelled out.

The reduction suggests that causal relations are to be analysed in terms of the relational properties of possible worlds insofar as they resemble each other. Now to keep the reduction honest, so to speak, the respects of resemblance must be with regard to the particular matters of fact upon which the reductionist supposes the causal relational properties of a world supervene. There must exist a pure factual level of description of the things in a world which subtracts out any "logical implications about their powers" (Ayer, 1972:115). However, it is far from

(3.29) "...clear that one can make sense of the idea of subtracting out from a property all 'logical implications about their powers'. Consider any ordinary property such as the property of being blue, or having a mass of 73 grams. Now try subtracting out of it, not just a particular causal power associated with the property..., but *all* causal powers. I think that such thought experiments about examples suggest what is persuasively argued on more general grounds, that 'what makes a property the property it is, what determines its identity, is its potential for contributing to the causal powers of the things which have it' [Shoemaker, 1980:114]. On this conception of properties, if we abstract away from the causal consequences of a property, there will be nothing left. The levels at which we describe the world are causal all the way down." (Stalnaker, 1984:159)

Strong grounds would appear to exist, then, for treating the causal relation as a semantic primitive, as an objective feature of the world, along with individuals. Since the Humean view was derived from an outmoded semantic theory this resurrection into respectability seems long overdue.

3.2.8 Localism vs Globalism

Certain metaphysical observations serve to connect some of the ideas in Stalnaker (1984) and Cartwright (1983), although once the connection has been made it will highlight a contrast. Van Fraassen (1980) argues for an anti-realist philosophy of science. In so doing he wishes to banish counterfactuals as having no part in a proper description of the way the world is. His argument relies on the previously noted context-dependency of counterfactuals:

(3.30) "The hope that the study of counterfactuals might elucidate science is quite mistaken: scientific propositions are not context-dependent in any essential way, so if counterfactual conditionals are, then science neither contains nor implies counterfactuals." (van Fraassen, 1980:118)

In arguing against this claim Stalnaker observes that:

(3.31) "...the claim that scientific statements are never context-dependent seems to me questionable" (Stalnaker, 1984:150)

He goes on to claim that scientific practice must at the very least

(3.32) "...provide a context for the interpretation of the language it uses to describe the world." (ibid.)

From Cartwright (1983) an even more radical claim with regard to the context-dependence of scientific propositions can be added. Insofar as the most central propositions of science: scientific laws, are true, they are only true, *ceteris paribus* (strictly: "other things being equal", but as Cartwright (1983:45) observes, more accurately "other things being *right*"). She derives this conclusion not from the usual "folk" science used in philosophical examples but from an examination of the fundamental laws of physics. For example, Snell's law, that the angle of incidence = the angle of refraction only holds for isotropic mediums. Only theoretical entities, and the complex and *localised* laws which describe them, can be treated realistically, but the simple unifying laws of basic theory cannot. So, insofar as fundamental laws can be treated as true descriptions of the world, they are radically context sensitive. This kind of metaphysical local realism with respect to causal laws, emergent in the philosophy of science, is just the world view which situation semantics adopts and is attempting to capture formally. In this light situation semantics could be put forward as the first attempt to trace the consequences for semantics and therefore most of contemporary philosophy, from the adoption of a new local realist conception of the world.

The obvious term of contrast for a localist conception is a *globalist* conception of the world. Localism has been tied to context sensitivity and perhaps, therefore, is intuitively clear. However, subsequently, although not explicitly, much weight will be placed on a contrast between localist and globalist conceptions. Characterising the distinction can be best achieved by a quote from Ian Hacking (1983). At the end of a chapter supportive of Cartwright's views, he introduces the following evocative picture:

(3.33) "God did not write a Book of Nature of the sort that the old Europeans imagined. He wrote a Borgesian library, each book of which is as brief as possible, yet each book of which is inconsistent with every other. No book is redundant. For every book, there is some humanly accessible bit of Nature such that that book, and no other, makes possible the comprehension, prediction and influencing of what is going on. Far from being untidy, this is New World Leibnizianism. Leibniz said that God chose a world which maximised the variety of phenomena while choosing the simplest laws. Exactly so: but the best way to maximise phenomena and have the simplest laws is to have laws inconsistent with each other, each applying to this or that but none applying to all." (Hacking, 1983:219)

A globalist might adopt the following view. Once the simple unifying laws of science are specified and the initial conditions surrounding the origins of the universe identified, then the dynamics specified by those laws will run without exception predicting all that has happened and all that will happen in the future. In contrast the localist will always being asking what state is *this part* of the world in *now*. The initial disposition of the universe may involve many different states all of which may require different mutually inconsistent dynamic laws to apply. Once they have applied a new local state will emerge which may require the application of laws which are inconsistent with the laws that lead to the present state.

In predicting such a world an organisms inferential mechanisms, which attempt to track reality, will have to be sensitive to changes in the circumstances of an inference. People may maintain laws which are mutually inconsistent but which have their own appropriate domains of application. This is a surprising contention when related to the laws of physics but it is surely the norm regarding social conventions. For example, at club A you may have to wear a tie to get in, whereas at club B you may have to not be wearing jeans, at club C you may have to wear jeans, at club D...etc. However, there is no unifying higher level generalisation which truly describes what happens at all clubs. There is a general heuristic to the effect that most clubs have entrance restrictions relating to dress, so its probably a good idea to check if any are in force at the club your going to. However, this heuristic is not part of the *true* description of the social conventions in operation. This

argument mirrors Cartwright's against a realistic interpretation of the fundamental unifying theoretical laws of physics. These laws play a unifying role but almost in virtue of this fact they cease to contact reality. To see how high level theory can be applied to any particular case requires various mathematical simplifications and approximations. Nothing in the theory validates those approximations, but if anything is true it is these particular *phenomenological* laws connecting theory to the phenomena. The same piece of theory may connect to different phenomena via different simplifications and approximations. This may lead to different and *inconsistent* phenomenological laws.

As described in Cartwright (1983:19) and quoted in Hacking (1983:219), the globalist picture of the human mind echoes Duhem's vision of the minds of French physicists while the localist picture echoes his vision of the mind of an English physicist:

> (3.34) "The French mind sees things in an elegant, unified way. It takes Newton's three laws of motion and the law of gravitation and turns them into the beautiful abstract mathematics of Langrangian mathematics. The English mind, says Duhem, is in exact contrast. It engineers bits of gears, and pulleys, and keeps the strings from tangling up. It holds a thousand different details all at once, without imposing much abstract order or organisation."

3.3 Below "causal" constraints and attunement

Stalnaker withholds from abandoning possible worlds because he still retains a moderate empiricism, whereby although he accepts that *some* methodological policies are causally grounded, *most* probably are not. On his conceptualist interpretation, he can still handle the latter using the Humean projection strategy. What, though, can the situation semanticist say about, what will be called *below causal* constraints? Two main issues are raised by constraints which don't obviously appear to be part of the ultimate causal structure of the physical world. First, how they are "grounded"? For causal constraints, the assumption is that they are, at some level, grounded in the physical make up of the world and ultimately possess a description in the vocabulary of physics or the special sciences. However, other constraints involving dispositions, promises, moral obligations etc. do not seem similarly grounded. Second, how do people acquire or become attuned to these constraints? For the causal case, enumerative experience of the instantiations of causal laws seems the most obvious mechanism, but could this work for, say a conditional promise?. A further problem concerns the possibility of error. It would appear that the causal constraints which hold between states of affairs in the world and mental states (cf. diag. 2.1), may not leave room

for being mistaken. This problem will be looked at first.

3.3.1 Error and projectibility

In the section "Inference and information gain" (cf. 2.4) constraints holding between mental states and the world were treated as *bona fide* constraints. This circumvents early objections to a semantic theory which takes the meaning of an utterance, eg. "There is smoke" to be a constraint holding between the utterance type ("T") and situation types where there is smoke (T), C: "T" ==> T. In some circumstances T may not be actual which creates the problem of error (Dretske, 1985). The problem of error involves one of the relata of a constraint not being actual. It applies equally to erroneous mental states and is created by the situation in diagram 3.1: *a* is in mental state ΨT(1)'', where the content of this mental state reflects *a*'s awareness (not necessarily conscious) of the environmental circumstances. His further mental state ΨT, is ambiguous between T': there is smoke, or T: there is dry ice.

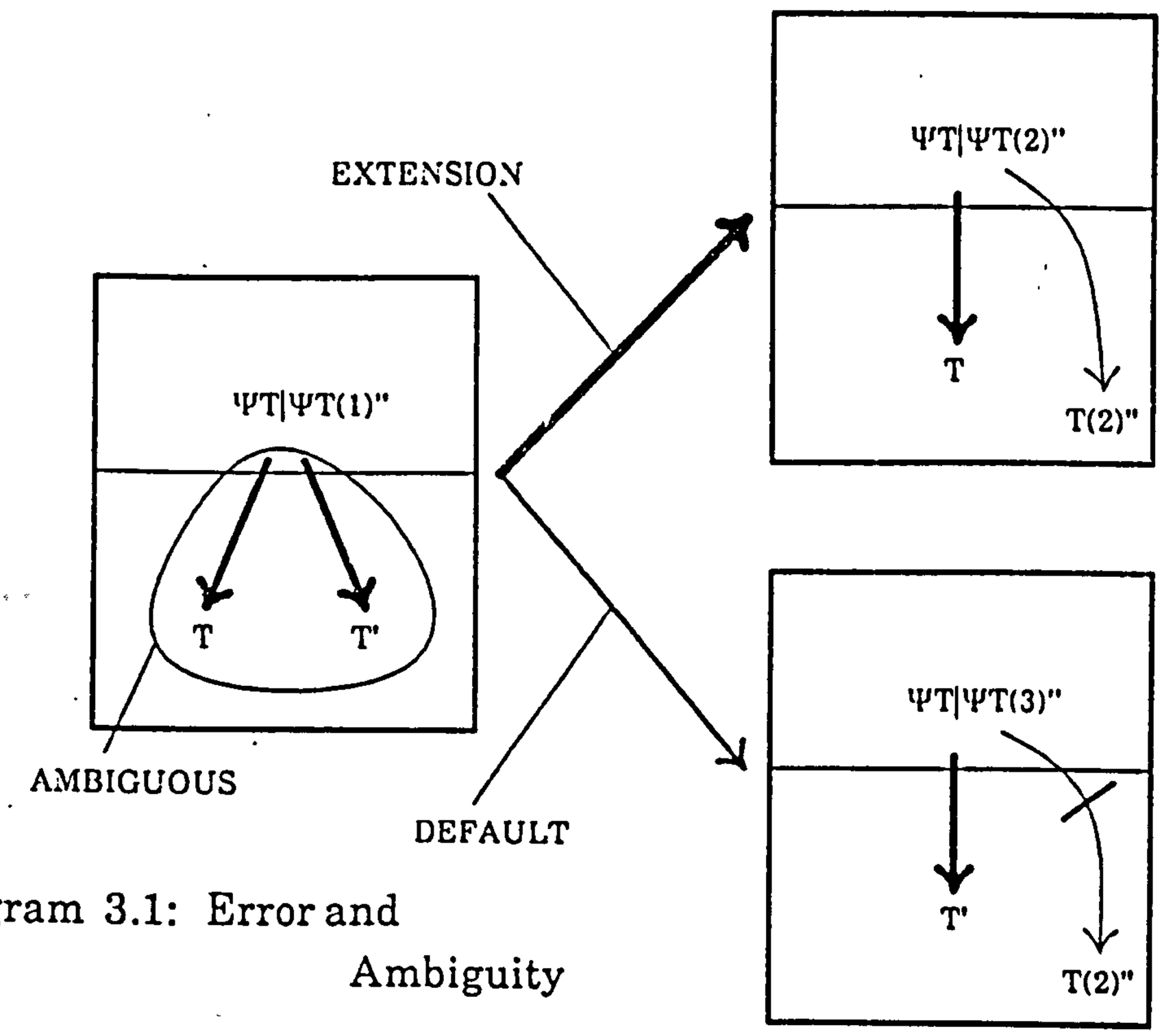

Diagram 3.1: Error and Ambiguity

ΨT(1)'' is insufficient to decide between these states of the world. If the information is then extended such that *a* is in mental state ΨT(2)'' the content of which is that he was on candid camera, then mental state ΨT now unambiguously indicates that T. There are really two constraints: (i) ΨT ==> T1 T(3)'', and (ii) ΨT ==> Tl T(2)''. Mental state ΨT(1)'' *a* is ambiguous between ΨT(2)'' and Ψ T(3)''. *a* may nonetheless *default* to T(3)'' from ΨT(1)'', ie. the most familiar context, to derive the interpretation T (or perhaps survival related due to other constraints to which *a* is attuned). Error is possible because the relation between a word or mental state and the objects and relations they denote is naturalised in a way that mirrors the context dependent relations between other states of affairs.

The meaning of, say "green", can be conceived of as a causal *relation* between objects in the world which reflect light of thus and so wavelengths and the effects that that reflected light has on an organism with thus and so perceptual equipment. As long as, in the appropriate circumstances, objects which reflect light of around 510 nm tend to produce the same effects in organisms like us then the predicate annexed to that relation will be projectible. Both the intrinsic properties of the object and peoples' specific cognitive equipment enter into the meaning of the term. "Green" does not attach to the sensation produced by objects which reflect these wavelengths. Rather, the causal role of these mental states (caused by these objects) in the whole functional economy of constraints to which a person is attuned determines the internal meaning of "green", ie. its cognitive significance. A predicate like "grue" would involve that relation changing such that light at these wavelengths produces different effects after the year 2000 a.d. This is sufficient to maintain an asymmetry between projectible predicates like "green" and non-projectible predicates like "grue".

3.3.2 Dispositions and conditional promises

Some examples will first be introduced. Each example is a particular indicative conditional which suggests a particular methodological policy or constraint for dealing with information about different domains of human activity. In each case it will be shown that each can be used to license counterfactuals and, therefore, each needs to be treated as fact stating on the assumption that the corresponding counterfactual is true. Therefore, an issue arises concerning how they are actually grounded.

Dispositions

(3.35) If I go to work, I always buy a newspaper.

(3.35′) My wife contemplating her lack of a newspaper and my laziness:

If he had gone to work, he would have got a newspaper (...and I would not have to go and get one).

Conditional-Promises

(3.36) If you cook tonight, I will wash up for a week.

(3.36′) To my wife a few days later; she is complaining about it being her turn to wash up again:

If you had cooked a few days ago, I would have washed up all week.

The situation created by these examples is diagrammed in Diag. 3.2(a). At the end of chapter 2, it was argued that a clear distinction should be retained between mental state transitions and the constraints upon which they are grounded. To keep this distinction clear the symbol "-->" will be used in the text to describe transitions between mental states. However, as we will see, it can be the case that one persons state transitions are ulitmately grounded upon another individuals action guiding state transitions or *dispositions*. In which case, the first individual's disposition provides the grounding for the second individual's state transitions or habit of inference. In diagram 3.2(a), the state transition ΨS: ΨT --> ΨT′ (the background types "| T′" will be left implicit), and the constraints C′: ΨT ==> T, and C′′: ΨT′ ==> T′ are all actual, but it is questionable whether the constraint C: T ==> T′ is also actual. Diag. 3.2(b) illustrates the projection strategy for dealing with this case suggested by Stalnaker. The twin arrows are the lines of *projection* on to the world. The situation in 3.2(b) is the Humean one, ie. there is no relational structure in the world. For Ryle, all we can talk of for philosophical purposes are T and T′. All the attempted analytic reductions looked at have tried to retain this picture while accounting for relational structure as supervenient on the intrinsic properties of the relata. Stalnaker on the other hand falls half way between the situation in Diag. 2.1 and in Diag. 3.2(b).

Examples (3.35) and (3.36) are relatively unproblematic, however, they do raise some interesting issues. For (3.35) the story that has been told so far can not hold directly. Suppose you regularly observe me buying a newspaper whenever I go to work. You therefore develop the disposition to infer that I will stop for a newspaper on discovering I am on my way to work. Let T: I go to work, and T′: I buy a newspaper, then your disposition is described by the state transition ΨS_{you}: ΨT_{you} --> $\Psi T'_{you}$ being actual. However, no C: T ==> T′ is actual. Rather your disposition ΨS_{you} is *grounded* on my disposition to do T′ when I do T, ie. my disposition: ΨS_{me}: ΨT_{me} --> $\Psi T'_{me}$, is actual. So your disposition is

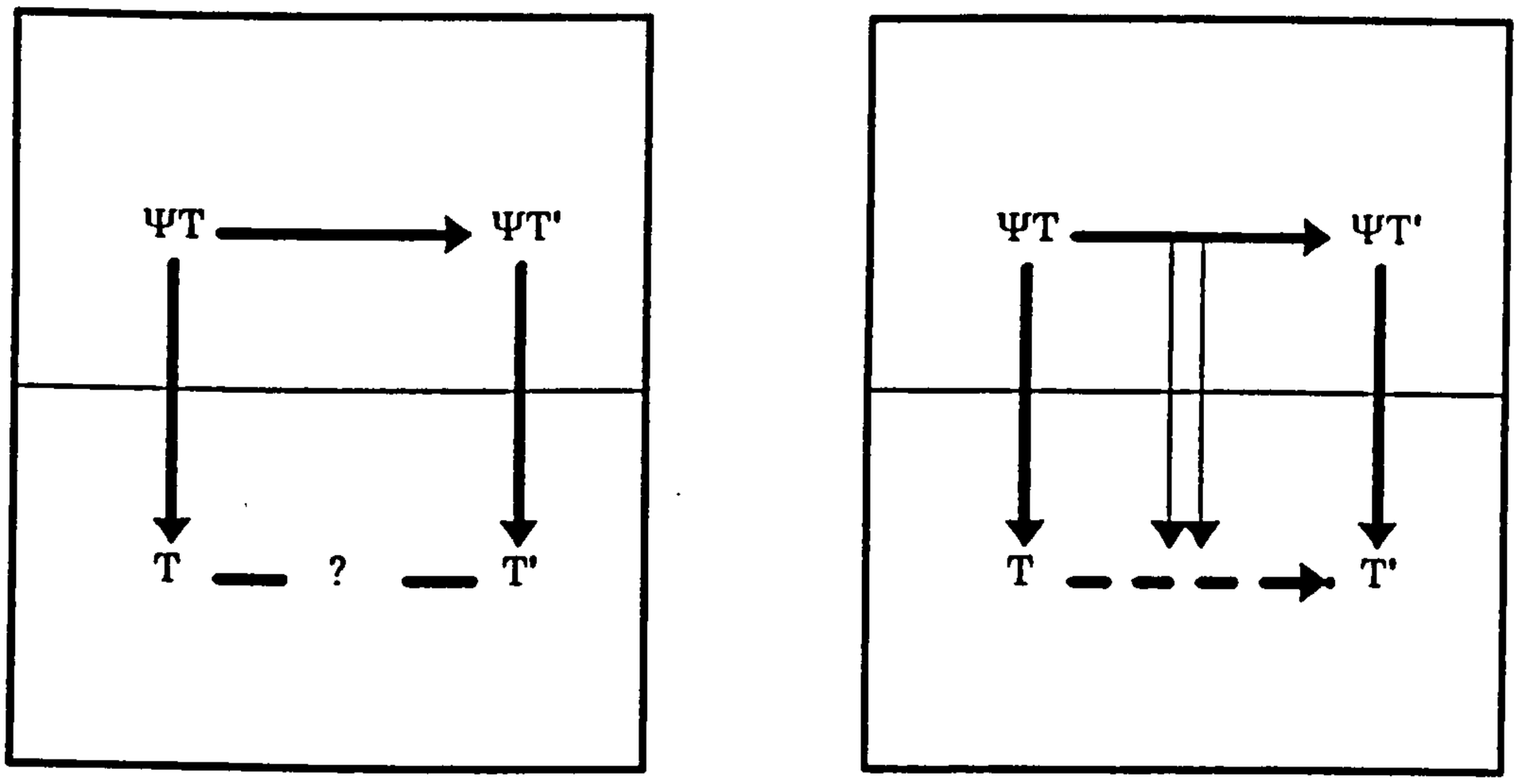

Diagram 3.2: Grounding below causal constraints and the projection strategy

grounded in an actual relation, ie. my disposition ΨC_{me}. The picture now looks like that in Diag. 3.3. So, there is a causal chain connecting T and T′, but it is mediated by my mental states and dispositions. This situation is no different from that in science, ie. unobserved (or unobservable) underlying *causal* mechanisms (referred to as "microscopic") are postulated which explain observed macroscopic regularities. As cognitive scientists, one may quibble over their precise computational nature, but one is unapologetically realist about the existence of certain inferential dispositions (cf. Stalnaker's observations on Dummett's (1978) assessment of the realism and reduction problem. Pace Dummett, proffering a reduction, say to some sequence of computations, does not immediately imply anti-realism with regard to dispositions).

However, there is a discontinuity between your disposition ΨS_{you} and mine, ΨS_{me}, with respect to how they were acquired. In section 2.4, attunement was treated as a state, ie. the state in which a transition between mental states (which commutes with an actual constraint in the environment) was actual. But how do people become so attuned? Is it bottom up and inductive or is it a process involving more active top-down cogitation? The example may

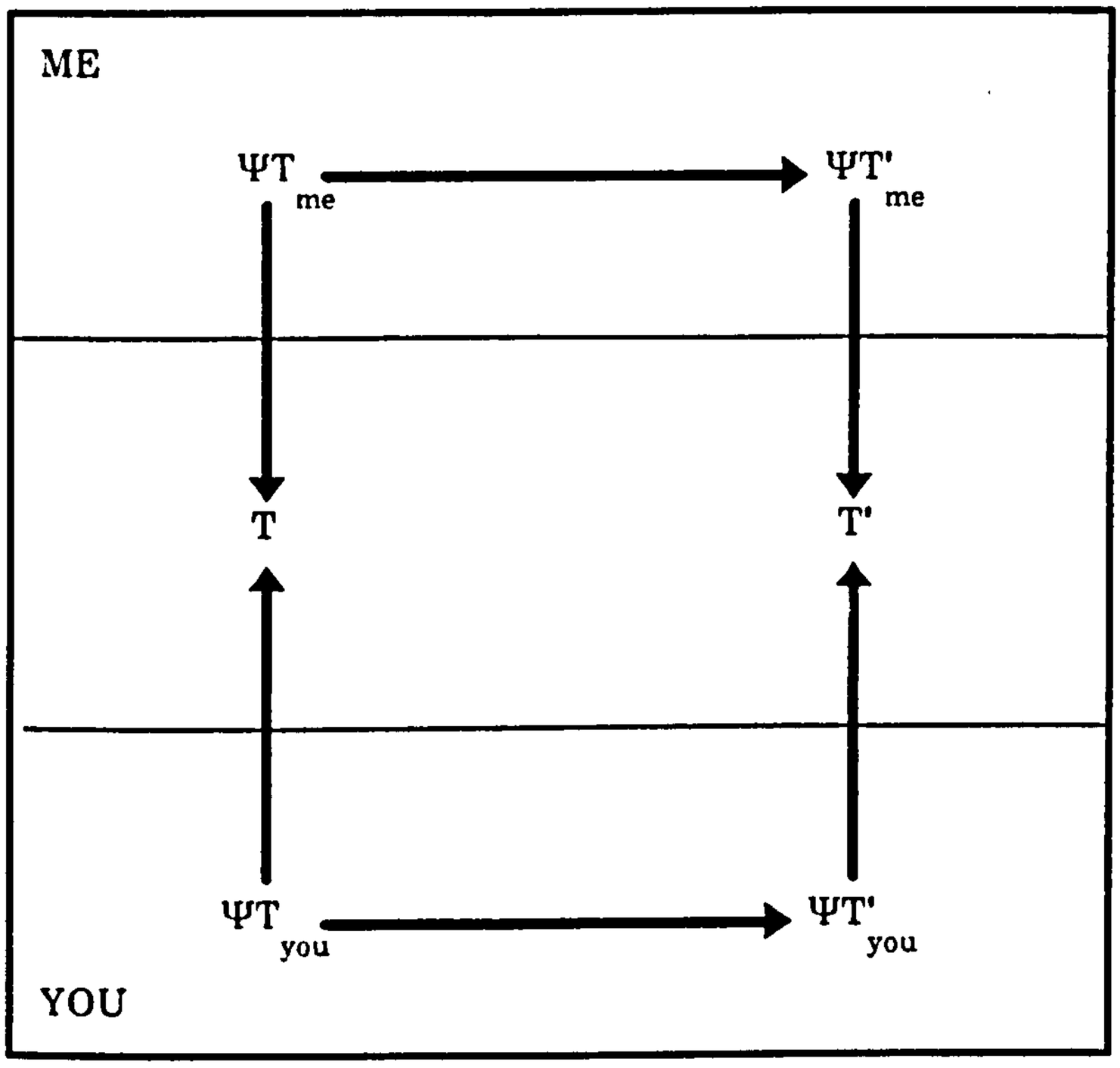

Diagram 3.3: Grounding Dispositions

help answer some of these questions. First, it is fair to say that the received view is that getting attuned is some form of inductive process. However, such a process could only account for ΨS_{you}, ie. your observations of my behaviour have attuned you to ΨS_{me}. But my possession of ΨS_{me} can not be the result of my observations of my own behaviour! At some time I must have made the conscious decision to stop for a newspaper. As I make this conscious decision more regularly it becomes sedimented into an automatic habit, just like learning to ride a bike. However, the element of conscious decision making can not be excised from an explanation of the process. The same may be said in your case: ie. it was your conscious decision to attend to my behaviour. But is this necessary? In ones perceptually guided action in the world one can not always decide what to attend to and what not to attend to. You may only realise your attunement to ΨS_{me} when you feel surprise at my walking past the paper shop this morning. Examples like this tend to argue for an active top-down component as well as bottom up processes in people getting attuned to constraints. (As observed above, Hume also erred in his over emphasis on passive bottom-up enumerative procedures as the primary source of habits of inference).

Example (3.36) poses new, but related problems. Suppose my wife believes that my promise to wash up for a week if she cooks tonight is sincere, ie. a few days later, she is inclined to believe that (3.36ˆ) is true. What process could be responsible for (i) my possession of this intentionally imposed constraint on my behaviour, and (ii) my wife's belief that I possess it, which inclines her to believe (3.36ˆ)? No habit forming process, involving active cogitation or not, has taken place. But it would appear that all relevant conditions concerning how we both handle information regarding this constraint are satisfied, ie. in our social dyad our behaviours are such that it would appear we are both attuned to this constraint. This seems to argue for essentially "one-shot" processes of entering in to a state of attunement. No habit forming forming process either initiated by active decision making or from the environment has occurred. The active decision stands alone as the source of the attunement. Promising is simply another way (process) by which people can enter into a state of attunement to a constraint. This constraint is one of, what I will call, my *fleeting* dispositions, ie. the antecedent is tied to a specific space-time location. (Cf. above it was noted that not *all* constraints are general, some, like (3.36) can be specific, ie. they are tied to a specific space-time location). However, despite its specificity, when it comes to grounding the relation it causes no more problems than (3.35). If a visitor asked why I washed up all week, the reply: "I promised my wife that if she cooked the other night I would wash up all week" is appropriate and amounts to the assertion that this constraint, ie. my intentionally imposed disposition, is (was) actual.

Not all bottom-up inductive factors are excised from this account, however. My wife's belief in my sincerity, is most obviously based on inductions from my past record of keeping promises. Moreover, active cogitation is probably even more implicated than at first glance. If my wife did cook that night, then my commitment to wash up that week may be based not only on my decision to commit myself to this disposition but also on my belief that one *ought* to keep ones promises. That is, the enforcement of this constraint in our social dyad is in part mediated by our joint assension to various moral obligations we believe to be in force. In other words it may rely on other types of constraint to which we are mutually attuned. This leads naturally to the next set of examples.

3.3.3 Moral obligations and social conventions

The same policy for introducing examples is followed as in the last section.

Moral-Obligations

(3.37) If you borrow money, you ought to pay it back.

(3.37′) If you had not borrowed the money, you would not have to pay it back.

Social-Conventions

(3.38) If you enter this club, you must wear a tie.

(3.38′) Replying to my friend's wife concerning his whereabouts, in the knowledge that he was last seen tie-less and she think's he has gone to the club:

If he had entered the club, he would have been wearing a tie.

The issue (3.37) and (3.38) raise is whether they can (or indeed, should) be treated in the same way as (3.35) and (3.36). In the case of both (3.35) and (3.36) the picture is as in Diag. 3.3. Although, there is no direct causal connection between T and T′, there is an indirect one via the dispositions of other cognitive agents. Both (3.35) and (3.36) were to do with the *contingent* behaviour of *individuals*. However, examples (3.37) and (3.38) concern *group* behaviour about which there appears to be some independent *necessitation*. This does not appear to fit Diag. 3.3. Take (3.37), and translate it in to the terminology used to discuss (3.35). It seems most natural to say that C: T ==> T′ is actual. The reasoning is as follows. It could be the case that (3.37) is explained as a generalisation of (3.35). Your attunement ΨS_{you}: ΨT_{you}[borrow money] --> $\Psi T'_{you}$[pay it back], is based not just on observing me doing so, which is then grounded in my corresponding disposition ΨS_{me} being actual, but on your observation of lots of other people also apparently being similarly attuned. So, the reason you find yourself compelled to repay loans is due to your attunement to the fact that most other people you observe do so. However, this clearly has an air of circularity about it. Dretske (1985:12) observes, it,

(3.39) "...is like trying to understand morality [moral constraints] in terms of the 'constrained' behaviour of good people. Such behaviour is constrained, I suppose, but the interesting question is what constrains it."

However, just this form of circularity was invoked by Goodman in the *entrenchment* response to his new riddle of induction. On his pragmatic philosophy, such circularity can be tolerated. Perhaps this is because there are no specific *institutions* which prescribe certain predicates and proscribe others with respect to their projectibility (although scientific institutions may be a candidate). However, in Society there are institutions which prescribe certain behaviours and proscribe others with respect to their moral rectitude. And it is the rules they lay down which ultimately ground moral obligations, ie. there is a C: T ==> T′.

(Note that the problems created by (3.36) can now be closed down. If the rule is on the club's books then thats that.) However, the question Dretske is raising is why do we feel compelled to be constrained by these moral prescriptions?

Grounding is not the problem. Various moral obligations are underwritten by institutions which also enforce them. Why we feel constrained to obey them seems to be a question that can also be asked of physically grounded constraints. For causal constraints the answer is obvious. Relative to physical survival it is adaptationally advantageous to allow ones behaviour to be constrained by the physical structure of the world. Adhering to these constraints has associated benefits, eg. continued survival, and associated costs for ignoring them, eg. death or injury. However, as our physical survival becomes dependent upon societies, so the need for social survival drives the need for restrictions on peoples individual behaviour. Relative to social survival it is adaptationally advantageous to allow ones behaviour to be constrained by the conventions of that society. Adhering to these constraints has associated benefits, eg. approval, emotional stability etc., and associated costs for ignoring them, eg. prison, emotional instability. Physical danger, and fear of not conforming tend to have similar emotional, and physiological effects. It thus seems that the mechanisms responsible for people's compulsion to conform to the laws of society are likely to be similar to those responsible for peoples compulsion to conform to the laws of nature.

That the constraints imposed by societies are properly grounded is going to make the same differences as those imposed by physical constraints being properly grounded. Cartwright (1983; cf. above) argues that the difference actual causal laws make is whether peoples goal directed strategies are effective or not. For example, stopping smoking is an effective strategy for prolonging life only if smoking is causally responsible for some life shortening ailment, in this case lung cancer. Similarly, wearing a tie is an effective strategy for getting into the club only if it is a rule that I must wear a tie to get in.

Learning what is and what is not an effective strategy can be achieved in similar ways. However, things can go wrong. For example, I may be interested in building a lighter than air craft. If I believed that burning increased the weight of a compound substance because burning involves the substance losing a constituent of the compound which has negative weight (ie. I adopt the phlogiston theory) then I may be inclined to get some of it into a balloon. I may do this by burning a substance underneath the balloon and observe, consistent with my theory, that it floats. However, I then decide to collect and store some and find out that the cooled negative weight substance no longer makes my balloon float. So to

construct plans of action it is important to be in possession of the right predictive constraints. But discovering what they are may only turn up in the long run as an active process of using them to make predictions. Similarly, I may be interested in getting into the club. I may pass the club regularly and observe everybody going in is wearing a tie. So, I go along one night wearing a tie expecting to get in but I am refused entry. Upon *asking the question* why? I am told that I am wearing jeans, and your not allowed in wearing jeans. So there is an asymmetry. In the case of the social convention perhaps the best way of discovering an effective strategy is to ask someone. Things are not so straightforward with the physical world.

3.4 Induction and information gain

Adopting a position on the nature of law-like relations is not independent of how people come to acquire them. In 3.2.4 it was stated as a putative desiderata that the epistemological issue of induction and the semantic issue of the meaning of causal statements should be kept separate. However, they are intimately related. Particular positions on the semantics of these statements condition the kind of answers obtained concerning how they come to be known. The role of causal laws in providing explanations constitutes another area where the semantic choices made have ramifications (Hempel, 1965; Bromberger, 1965; Salmon, 1971; van Fraassen, 1980; Cartwright, 1983). Goodman's *Fact, Fiction and Forecast* summarise's many of the objections to a theory of confirmation based on the semantics of the material conditional and the idea that confirmation is deduction in reverse. However, this is not the place to explicate a full theory of explanation and confirmation as a rival to alternative philosophical positions. Rather the aim of this section will be to outline the conception of the inductive process underwritten by the view of constraints as the context dependent laws that allow information gain. As argued in the last section, no essential difference will be admitted between causal constraints and *below* causal constraints. The same implied mechanisms apply to both cases.

3.4.1 The predictive cycle

The best way to outline the process is by introducing a diagram. Diagram 3.4 is a generalisation of diagram 2.1. Let us be absolutely clear concerning the function of the diagram. First, of all what it is not. It is not supposed to function as explication of precisely how people become attuned to predictive, information gaining constraints. Rather its function is

to elucidate the processes involved in determining the *circumstances* under which a constraint to which someone is attuned allows information gain or, in other words, is predictively successful. All constraints are *context dependent.* The primary goal of induction is to establish predictively successful rules. Therefore, the processes which require explanation are how people come to identify their appropriate domains of application wherein the predictions made are invariably successful. The distinction being cleaved to is familiar from the philosophy of science and holds between the *context of discovery* and the *context of justification.* The purpose of the diagram is to explicate the processes of justification once a hypothesis about a predictive constraint has been discovered modulo the context dependence of all such constraints. It will be found that procedures like falsification turn out to be maladaptive when the problem is cast in this way. However, a specification of a competence model of human inductive abilities would be incomplete without an outline of the processes of discovery. Four such processes were introduced in section 3.3. They are listed below:

1. Bottom-up enumerative learning.

2. Top-down enumerative learning.

3. Bottom-up "one shot" learning.

4. Top-down "one shot" learning.

1 and 2 concern the repetitive and largely unconscious processes involved in acquiring a habit. In the case of 1, repeated observations of a regularity in the environment may lead to an attunement to a predictive constraint. Someone may only become aware of their attunement to such a regularity when it fails. 2 involves the way in which repeated conscious decision making may become sedimented into an unconscious habit. 2 and 3 concern the processes by which a constraint is suggested by a single salient experience. In the case of 3, this may be given directly by the environment, ie. a particularly salient observation of a co-occurrence or sequence of events may lead someone to hypothesize that they always co-occur or one always follows the other. 4 involves the way in which simply being told that there is a rule in force, perhaps in response to a question or in a promise, can lead someone to use the rule to gain information.

The purpose of inductive procedures is primarily to establish the circumstances in which a rule permits successful prediction. Organisms must predict their environment in order to survive in it. A corollary to this process concerns the fact that only those transitions between mental states which are grounded in an actual constraint will turn out in the long run to be predictively successful. Upon discovering what seems a plausible constraint, the

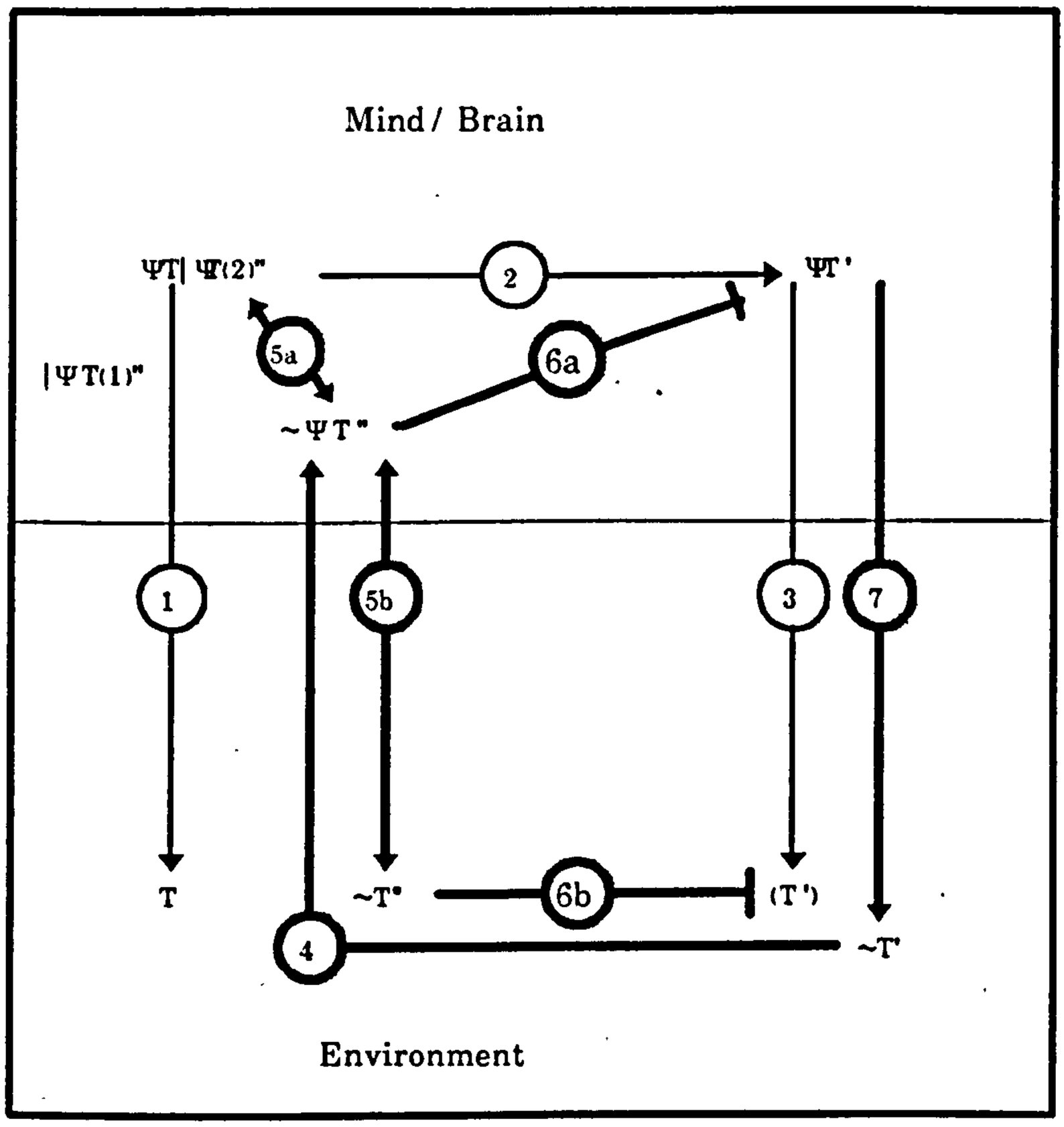

Diagram 3.4: The Predictive Cycle: Response to Predictive Failure and Constructing the Default Hierarchy.

best way to discover its domain of application is to use it in making predictions. Diagram 2.1 represents one predictive loop in which the prediction was successful, *a* gained information. Diagram 3.4 shows this cycle and what happens subsequently when a prediction fails. Response to predictive failure is the principle method of isolating the appropriate circumstances in which a constraint is informational. The arcs in the diagram are labelled numerically to indicate the processing stage involved.

The narrow arcs indicate the process up to predictive failure, the broad arcs indicate the subsequent response. At stage 1, *a* is in state ΨT which in the appropriate circumstances

ΨT(1)'', involves T. Given possibly overlapping circumstances, ΨT(2)'', which *a* must assume *ceteris paribus* (interpreted a la Cartwright as "all other things being *right*") *a* transits to mental state ΨT' [stage 2]. This leads to the expectation of T' (brackets indicate an expectation) [stage 3]. However, T' is not observed. This requires *a* to determine what additional condition could be in operation which could have precluded [arc 6] T' despite T being actual. The hypothetical mental state corresponding to this condition is marked as $\neg\Psi T_1$'', the negation sign indicates it is the dual of this background type which could preclude ΨT'. $\neg\Psi T_1$'' could arise from any direction marked by the bi-directional stage 5 arcs. The state transition ΨS: ΨT --> ΨT' is conditional. Prior beliefs concerning ΨT(2)'', could function to suggest $\neg\Psi T_1$'', which could have blocked the transition to ΨT', and therefore the expectation that T' is actual. In which case $\neg\Psi T_1$'' will derive from the background type ΨT(2)'' [stage 5(a)/down]. This will suggest the hypothesis that $\neg\Psi T_1$'' is actual [stage 5(b)/down]. Alternatively, *a* may be unaware of any relevant conditions which could block ΨS. In this instance some salient feature of the environment $\neg T_1$'', may have been present when *a* observed T. This may be lead to mental state $\neg\Psi T_1$'' [stage 5(b)/up]. $\neg\Psi T_1$'' is hypothesized to preclude ΨT' [stage 6(a)/(b)], and could be encoded in the background type of ΨS [stage 5(a)/up]. So, despite T being actual, because of $\neg T_1$'', T' was not, thus re-establishing the predictive completeness of the cycle [stage 7].

The predictive failure could have come about for two other reasons. First, *a* was mistaken about T being actual. When *a* makes the assumption that ¬T' is good evidence for his falling into perceptual error is the only time that an inference corresponding to modus tollens would appear rational. The only other grounds for T' not being actual are that the constraint C is itself not actual. However, this is unlikely to be the preferred inference. The discovery procedures by which *a* came to be using ΨS in an information gaining way have already given some independent reasons to trust that C is actual. In the light of the context dependency of predictive constraints, *a* is unlikely to give up ΨS on the basis of *one* predictive failure.

The process outlined in diag. 3.4 will now be exampled using a taxonomic and a non taxonomic constraint. First of all the appropriate mechanism of ignorance needs to be identified, ie. the factor which demands an inference in order to resolve ignorance. For non-taxonomic, dynamic constraints, time is the appropriate mechanism of ignorance, an organism is attempting to predict an unknown future. For taxonomic, static constraints, the appropriate mechanisms are either language or space. An organism is either making inferences from an incomplete description or inferences about incomplete knowledge of a perceptual event which could be resolved by investigatory actions in space-time.

Taxonomic: An interlocuter informs you that tweety is a bird ("T"). Due to the conventional constraint between words and mental states you go into a mental state the content of which is that "tweety is a bird" [1]. Due to your attunement to the constraint that "birds fly", you make the default inference (ie. make a *ceteris paribus* assumption) to "tweety can fly" [2]. Which leads you to expect that tweety can fly [3]. However, you learn that tweety can not fly. This leads you to derive the default condition ($\neg\Psi T_1'$), that "tweety is a penguin" [4](it is implicit in your inference to "tweety can fly" that "tweety is *not* a penguin, hence the appositeness of marking this condition as the dual "$\neg$"). This could be because you have already discovered this default [5(a)/(b)/down], or because you now check tweety and discover that it is particularly salient feature of tweety that he is a penguin [5(b)/(a)/up]. Either way, tweety being a penguin could preclude tweety flying [6a,6b], which re-establishes the predictive completeness of the cycle [7].

Non-Taxonomic: At t_1 (0900) you are in mental state ΨT the content of which is that "Fred is passing your window"; this involves T being actual [1]. Due to your attunement to the constraint that Fred passing your window at 0900 means Fred is going to buy his morning paper, you make the default inference (ie. make a *ceteris paribus* assumption) to "Fred is going to buy his morning paper" [2]. Which leads you to expect that Fred is going to buy his morning paper [3]. However, you learn, at t_2, that Fred walked straight past the paper shop. This leads you to derive the default condition ($\neg\Psi T_1'$), that "Fred is trying to get to work early" [4]. This could be because you have already discovered this default [5(a)/(b)/down], or because you remember that it *was* particularly salient feature of Fred that he was on his bike which he only rides when he is in a hurry [5(b)/(a)/up]. Either way, Fred being in a hurry to get to work could preclude him stopping for a newspaper [6a,6b], which re-establishes the predictive completeness of the cycle [7].

Deriving the appropriate defaults in the background type functions as an answer to the question "Why $\neg T'$" on the assumption of T and that C is context sensitive. The answer to this question is the relevant default condition, $\neg T_1''$. The default condition functions as an *explanation* of why $\neg T'$ given T and that C is context sensitive. The explanation works because the relevant condition, $\neg T_1''$, *precludes* (is negatively causally related to) T'. Furthermore, if T' is actual, then a good explanation for this, ie. a good answer to the question "Why T'?", is that T, and the constraint C was actual. There is an asymmetry between reasoning in the predictive direction and the explanatory direction. It is adaptationally advantageous to be able to correctly predict the world. If one is to do so then it must be assumed that on learning T, all the circumstances are right to infer T', *even in ignorance of whether the assumption holds*. However, explanations are always *post hoc*, nothing

immediately depends on them, so the question concerning which conditions held and which didn't can be considered at leisure. Moreover, in assessing a constraint one will be concerned primarily with predictive success. The situation that doesn't matter is where there is a many - one relation between T_i and T′ (ie. many constraints lead to T′). But a one - many relation between T and T_i′ is predictively less advantageous as then T is ambiguous. However, the former situation will lead to explanatory ambiguity, ie. T′ may be explained by the occurrence of any T_i. In general, reasoning to the best explanation of T′ is a less reliable process given the primacy of successful prediction.

3.4.2 Are constraints falsifiable?

Can a hypothesis about a constraint ever be falsified?. Context sensitivity blocks a straightforward Popperian falsification strategy. An instance of T and ¬T′ does not falsify but drives the construction of the default hierarchy within the background type. Popper (1959) tended to undervalue the importance of the evidential basis given by the discovery of the constraint (cf. above). If it was the product of many observations of a regularity or you were told about the constraint by a particularly reliable individual you are unlikely to give it up on observing one predictive failure. And again, the adaptational importance of prediction means that the possession of any rule, however limited its domain, is preferable to no rule at all; with no predictive constraints to guide its behaviour an organism is an impotent pawn of its dynamic, changing environment.

The need to predict an ever changing world indicates that falsification of a constraint may occur but it is only likely after *a long run* of persistent predictive failures. This conclusion is a result of general considerations concerning the adaptational requirements of an organism relative to a changing environment. It may prove more convincing if the line of argument was also reflected in a conception of scientific method. Psychological work on conditional reasoning has taken its competence models from philosophical/logical work on scientific methodology. This work has been taken to underwrite the normativity of the competence models and thus provide the psychologist with a definition of error. The conception of scientific method which has driven the psychological work I will go on to consider in the next chapter, has been Popper's (1959). However, there have been many subsequent developments in the area of methodology (Kuhn, 1970; Lakatos, 1970; Feyerabend, 1975; Putnam, 1974; Hacking, 1983). Most of these developments have questioned the plausibility of falsification in the light of the historical evidence on scientific progress. One of the most recent developments is due to Hacking (1983). Hacking's work is closely related to

Cartwright's and to Peirce's *pragmatism.*

Philosophical arguments over method centre on two issues (i) scientific realism, and (ii) rationality. The issues are intimately related. Does scientific method allow inquirers to discover the true structure of reality, and does it provide a rational procedure for fixating beliefs (theories). The positivist position, "results from the conception that seeing is believing" (Hacking, 1983:63). Truth is a correspondence to an external reality which is given in perception. This leads to positivist anti-realism with regard to unobservable theoretical entities and, for familiar Humean reasons, causation. Hacking on the other hand, argues for a realist account of causation and theoretical entities. The existence of theoretical entities is not necessarily tied to their direct observability but rather to the *process of inquiry* wherein the causal effects of a theoretical entity are regularly observed and manipulated. Hacking argues that the conception of inquiry he advocates derives directly from the pragmatism of Peirce. For Peirce, reality is

> (3.40) "...that which, sooner or later, information and reasoning would finally result in, and which is therefore independent of you or me." (C. S. Peirce, 1868, *Some consequences of four incapacities*, quoted in Hacking, 1983:58)..

To fixate beliefs requires, "...a method which is internally self stabilizing, which acknowledges permanent fallibility and yet at the same time tends to settle down" (Hacking, 1983:59), "truth is whatever in the end results" (Hacking, 1983:61).

Inquiry is a dialectic process between theory (reasoning) and data (information). For Hacking the truth of a law or the existence of a theoretical entity is determined by peoples' ability to regularly use those laws or causal properties to actively change and manipulate the world. Or, by way of paraphrase, when they become part of our repertoire of *effective strategies* for achieving our goals. Making this determination can only be achieved in the long run of using rules/laws relating observables to observables, or unobservables to observables in making predictions, designing machines or generally in *intervening or acting* in our environment.

In consequence, Hacking (1983) emphasises the role of experimentation in science. Experiment is characterised, on the basis of many actual examples, as having a life independent of theory. On a falsificationist methodology the paradigm of experimentation is the *crucial* experiment, ie. an experiment which decides between competing hypotheses. A paradigm case of is the Michelson-Morley experiment which on Popper's interpretation decided against the absolute Newtonian view of space-time and for relativity theory. As Lakatos (1970) also observes, this constitutes a *post hoc* rationalisation of what Michelson and

Morley took themselves to be testing which was between two rival views (due to Stokes and Fresnel) of the behaviour of the aether near the earth's surface (Hacking, 1983:254-61). It was relative to the aether that absolute measures of the earth's velocity were to be obtained. Scientific experiments may have various functions of which acting as a crucial test may be one. However, they are hard to find in the actual historical record. Purported paradigm cases turn out to have had very different functions.

Crucial experiments are only possible modulo competing theoretical positions which make contrary predictions. However, Hacking observes that much experimentation proceeds independent of theory. A great deal of experiment involves establishing the precise circumstances under which certain effects are observable, or to establish more precise measures of physical or other constants. Relative to my subsequent psychological interests it would be an interesting turn around to offer a descriptive analysis of experimentation on the Four Card problem. The principle phenomena concerned the failure to observe a Popperian falsification strategy. However, this was *not taken to falsify a Popperian view of what subjects should do*. Rather further experiments were designed to test the domain of the phenomena. It would appear that work in this area proceeds with a meta-theory which is inconsistent with object theory under test. What researcher's subject's should do has not been taken to determine what the researcher should do.

Falsification would appear to be a product of the process of inquiry which can only emerge in the long run. Actual scientific experimentation is not always aimed at the attempt to falsify hypotheses, but rather, as in the model of the predictive cycle, is often aimed at establishing the appropriate domain of the phenomenon under investigation. People do not tend to actively falsify hypotheses but rather as their utility to achieving their goals diminishes, ie. they fail to be predictively useful, they get used less and therefore fail to be re-enforced as part of an organism's adaptive predictive repertoire. However, if an organism needs to act in novel domains then the situation may emerge where its existing repertoire of constraints ceases to be predictively useful. Thus an agent may need to acquire new constraints which may make contrary predictions to existing constraints. These will require testing to assess the relative utilities of existing constraints against newly discovered constraints which may well lead to *eliminative* falsificatory reasoning.

Summary

In this chapter a foundation for the concept of a constraint has been suggested which

locates this situation theoretic concept in both recent conceptions of scientific laws and method, and within the semantic literature on conditional logics. It was seen how various attempts to reduce causal laws to other less problematic concepts failed. In part this is due to arguments concerning the possible worlds semantics for the conditional provided by Stalnaker (1968, 1984). In providing a metaphysical grounding for his abstract semantics it was seen that the selection function is often grounded in the causal relation. It was also shown how, pace Stalnaker, other constraint types can also be given a similar grounding. Once a conception of natural law as local and real has been adopted, the possible consequences for scientific method were traced through, grounded in the work of Nancy Cartwright and Ian Hacking. This suggested the model of the inductive process embodied in the predictive cycle.

In the next chapter, the discussion of competence will turn to more psychological concerns. Some assumptions concerning the force of normative models of confirmation and partial interpretation will be re-considered. The chapter will begin by offering a classification of types of inference. This will serve to locate and contrast the present competence model with other normative theories employed in psychological research. Some of the distinctions drawn in chapter 2 will also be shown to be reflected in the psychological data on conditional reasoning.

Chapter 4: The Selection Task: Re-interpretation

4.1 Introduction

The role of the preceding chapters has been to motivate a competence model of inference and induction which can provide the rational basis for subjects observed patterns of reasoning. The restrictions of presentation have demanded that this model be developed without direct appeal to the psychological data which provided the focus for identifying which aspects of the normative literature could be instructive. In this chapter this imbalance will be redressed, thereby motivating some further distinctions in the way certain normative theories have been taken to provide or, more importantly, fail to provide a rational basis for subjects behaviour on conditional reasoning tasks. Two principle domains of competence will be looked at.

First, the attempts to formally define a confirmation relation. It will be argued that the paradoxes which beset confirmation theory are less paradoxical when the domain of a generalisation is restricted to a small finite set of objects. Experimental results which *prima facie* indicate subjects are falsifying may be interpretable in terms of the adoption of a strategy of *confirmation* which is appropriate to small surveyable domains. Second, the implications of partial interpretation will be discussed modulo the observation of a *defective truth table* in many conditional reasoning tasks. Subjects tend to treat false antecedent instances as irrelevant to the truth or falsity of a conditional. *Prima facie* the phenomenon of defective truth tables can be provided with a sound rational basis in partial interpretation. It will be suggested that the consequences of defective truth tables/partial interpretation have been seriously underestimated in the literature. A corollary to this discussion will include some comments on Johnson-Laird's (1983, 1986b) claim that there can be syntactically equivalent semantic schemes.

However, this chapter will begin by making explicit the kind of inferential behaviour which the competence models of the last two chapters are designed to capture. This will make precise the contrast between the treatment of conditionals which the present model adopts and the treatments which have usually been taken as normative with respect to the psychological data on conditional reasoning. The dependence of the present treatment on *content* will then be exampled and discussed. Some experimental data which can be provided with

a rational grounding directly within the present model will then be introduced. The reason being that this data bears most directly on the some of the central concepts of the competence models.

4.2 Inference: Deductive, Eductive and Inductive

There are several well established distinctions in which to frame the question concerning the status of an inference (Trusted, 1979). The first is between ampliative and non-ampliative inference. An ampliative inference *amplifies* or goes beyond what is logically entailed by the premises. Therefore, by definition, deductive inference is non-ampliative. However, this distinction, as is often pointed out, is an unfortunate one. It is frequently articulated in terms of information: an ampliative inference yields information, whereas an non-ampliative one does not. However, a complex chain of deductive inferences can surely be informative. It can reveal information not initially explicit in the premises. It will be allowed that deduction can be informative in this sense. Perhaps a better way to express the distinction is that a non-ampliative inference fails to go beyond what is logically explicit in the premises. Then "ampliative" can be used to describe inferences which permit information gain.

Normally ampliative inferences are further subdivided into *eductive* inferences those that reason from known particulars to unknown particulars, and *inductive* inferences, those that reason from particulars to generalisations. Again this may be an unfortunate distinction. Someone's warrant for predicting an unknown particular from a known particular is normally given by some general law which subsumes those particulars. So, the ability to make eductions depends upon prior inductions. The terminology has changed from the last chapter but basically the distinctions are the same: to make predictions about particular events (eductions) requires knowledge of general laws, ie. constraints (inductions). The model of the predictive cycle simply makes the relationship between eduction and induction explicit. Repeated eductions, ie. predictions of the unknown from the known, constitute the basis of induction ie. establishing the generalisation which provides the warrant for those eductions. Ampliative deductions however, do not allow either of these two forms of reasoning. Eductions are information gaining in a sense which goes beyond making explicit what is logically implicit in a statement. An eduction permits predictions about future uncertain events. Subsequently in discussing the information gaining role of a conditional describing a constraint, we will be concerned with information gain only in the eductive sense.

Could deduction capture the inferential mode of inference from particular to the general?. There is a rule of universal introduction (∀) in natural deduction schemes. Perhaps when it is valid to universally generalise then there is a deductive argument from the particular to the general. However, it is only valid to universally generalise when the particular in question does not depend on any non-logical properties. For example, if one argues about a particular triangle in order to establish a general conclusion about all triangles, then the argument, to be deductively valid, must not rely on any assumptions concerning the properties of that particular triangle (Thomason, 1970). Proofs by *mathematical* induction proceed in this way. If the argument only works for equilateral triangles the argument is not valid for all triangles. However, inductive inference is to argue from the instances of F*a* to the conclusion that ∀xFx when F*a* is provided on *non-logical* grounds. The rule of Universal introduction explicitly excludes such cases. Induction is an argument from what one might call *assumptive* particulars to generalisations.

Within inductive inference Trusted (1979) identifies three subcategories:

(4.1) Spontaneous inductions.
(4.2) Inductions arrived at by reflective common sense.
(4.3) Inductions arrived at through critical scientific study.

Both (4.1) and (4.2) may be corrected by the more critical methods of justification available in (4.3). This classification conforms to the distinctions drawn in the last chapter between discovery procedures (4.1, 4.2) and justificatory procedures (4.3). 4.1 corresponds to what was labelled "bottom-up" learning, and 4.2 to what was labelled "top-down" learning.

Constraints are the law like relations which permit *eductive* inferences from known particulars to unknown particulars in an uncertain world. The competence models which have typically been taken as normative with regards to the psychological data on how people fixate conditional beliefs have been based on falsification (Popper, 1959). This conception of belief fixation derives directly from the logic of the conditional and the universal quantifier. Reasons why this procedure fails have already been discussed in chapter 3 and a more precise explication of the reasons for its failure will be looked at later on in this chapter in discussing the implications of *partial interpretation*. However, the view that conditionals describe constraints serves to contrast the present competence model of human reasoning abilities with more conventional models. The contrast is highlighted by considering 3 sources of the inferential behaviour of a conditional. First, are the rules of inference, for example modus ponens and modus tollens. These syntactic rules allow various inferences based on the logical form of the sentence. Second, truth conditions, given the truth or falsity of the constituents of a complex proposition conclusions can be reached concerning the

truth or falsity of the whole proposition. Syntactic rules of inference and truth conditions govern the logical behaviour of the conditional. However there is a third view; in the real world people require rules which permit information gain, in the sense outlined above. They need to predict future events, ie. to make *eductions*. The core hypothesis of this thesis is that only once psychology is in possession of a competence model of the information gaining or eductive behaviour of conditional rules will a rational basis for subjects inferential behaviour be forthcoming. The transitions between mental states which are subsequently grounded in the operative constraints in the world are *eductive* inferences. Eductive inference relies on relations between content, ie. constraints, not logical relations. It is unsurprising, therefore, that how beliefs about these relations are fixated relies crucially on pragmatic content.

4.2.1 Content and information gain

Goodman's grounds for rejecting any attempt to formally derive a confirmation relation was based on an argument from the influence of content on induction (cf. chapter 3). Knowledge of which predicates are projectible is a crucial determinant of framing hypotheses which are capable of inductive support. The dependence of induction on content can also be illustrated by the fact that spontaneous inductions to particular hypotheses are not independent of pragmatic world knowledge. Formal attempts to define a confirmation relation was concerned with justificatory procedures subsequent to discovery (cf. next section). The idea was to define a relation between particular instances and a general hypothesis, such that once discovered only those instances which are deductive consequences had the potential to confirm (cf. below). However, how a hypothesis is initially framed relies on particular pragmatic world knowledge over and above whether the predicates are projectible. For example, logically Fa & Ga instances seem equally to suggest the hypothesis that:

(4.9) $\forall x(Gx \rightarrow Fx)$ (eg. All black things are ravens)

or the hypothesis that

(4.10) $\forall x(Fx \rightarrow Gx)$ (eg. All ravens are black)

But it would be absurd to suggest that observations of black ravens equally invite the spontaneous induction to all black things are ravens as well as all ravens are black. Knowledge of black non-ravens makes only the latter hypothesis reasonable. The asymmetry here is given partly (cf. below) by knowledge of the respective sizes of the domains specified by

the predicates.

(4.10) is a nomic taxonomic constraint:

T_{Σ}: <<*Raven*, **x**; *1*>>

$T'_{\Sigma'}$: <<*Black*, **x**; *1*>>

(4.11) <<[n_t]=>, T_{Σ}, $T'_{\Sigma'}$; *1*>>

The order of indicating type and indicated type encodes information about the relation between these predicates (unary-relations). The asymmetry observed between (4.9) and (4.10) also reflects the natural *subject/predicate* order. Although logically both relations function as predicates, *raven* functions naturally as an object identifier, and *black* as a qualifier. This is also a pragmatic function of the predicates which is reliant on content. There would appear to be a natural classification of predicates into those which identify particular classes of individuals and those which qualify the properties of those individuals which may be shared by many classes. This establishes the order of predicates in a hypothesis which allows predictions. If you know its a raven you can predict it will be black, but if you know its black you can't predict it will be a raven. The relation which holds between these classes in virtue of pragmatic world knowledge can only be *class inclusion*. The domain of objects identified in the indicating type cannot be smaller than the domain of the qualifier in the indicated type. The relation of class inclusion prototypically licenses inferences which accord with falsification. Central to this mode of reasoning is the concept of single "instances", or relative to a unifying relation like "traveling" (cf. chapter 2) single occurrences, possessing various properties.

In a taxonomic constraint the instance in question is identified in the indicating type and qualified in the indicated type. This is why the class inclusion relation is appropriate. However, in a non-taxonomic constraint discrete occurrences of events are related by a higher order relation, like cause, enablement, permission etc. (Although the possibility must be admitted that dependent on pragmatic world knowledge the relation which does exist may permit inferences which concur with those licensed by class inclusion.) In non-taxonomic constraints there is a discrete indicated type. For example,

T_{Σ}: <<*Turn_light_switch:* **x**; *1*>> & <<*Switch_A*, **x**; *1*>>

$T'_{\Sigma'}$: <<*Light_on*, **y**; *1*>> & <<*My_hall_light*, **y**; *1*>>

(4.12) <<[n]=>, T_{Σ}, $T'_{\Sigma'}$ $T''_{\Sigma''}$; *1*>>

(4.12) expresses the nomic constraint holding between turning switch A and my hall light coming on. $T''_{\Sigma''}$ will include background types like the electricity is not cut off, the bulb is working etc. T_{Σ} and $T'_{\Sigma'}$ are discrete events which are causally related. The asymmetry is in part mediated by pragmatic knowledge of this causal relation. However, unlike the taxonomic constraint it seems that the indicated type can also be *educted* from the indicating type. If the light is on I can educt to the switch A having been turned. However, the asymmetry is partly re-established by the causal/temporal asymmetry which means that when the eduction tracks causality it is called *prediction* and when it reverses this direction it is called *explanation*. This too is a function of pragmatic world knowledge.

Constraints license eductive inferences from known to unknown. In particular circumstances a constraint will license an eduction which reverses the temporal/causal order. For example a causal constrain can possess a corresponding *explanatory* constraint which reverses *real* order. Explanatory constraints, while licensing eductions, reverse real order and are parasitic upon the actuality of the causal case. You can explain an effect by its cause but not a cause by its effect (but cf. below on *teleological* explanation).

Conventions are slightly different. Two forms of a convention can be distinguished:

(AP) If ACTION, then PENALTY, and

(PA) If PRE-CONDITION, then ACTION

They are related insofar as a failure of PA is likely to transitively involve an AP, eg. driving a car (action) without a driving license (precondition) is likely to incur a stiff fine (penalty). *Prima facie* AP mirrors the causal case. The incursion of the penalty can be explained by the performance of the action, but the performance of the action can not be explained by the incursion of the penalty. *Unless* it was someone's intention to incur that penalty, ie. this was the *goal* he had in mind when performing the action. There is a distinction between *causal* explanation and *teleological* (goal directed) explanation. Relative to goal directed explanations the order of indicating and indicated types may be reversed. *However* this applies equally well to causal constraints which are employed in teleological explanations. For example, (it is assumed that an explanation is an answer to a why-question: cf. van Fraassen, 1980). If the question is posed, "Why did she whip him?", the explanation may be that she wanted to cause pain, ie. it invokes the causal relation between lacerated skin and pain. The pain is caused by the lacerated skin, although the explanation of her action was the effect, not the cause. Teleological explanations, reverse the "natural" ordering of events (actions) in time.

However, even goals are directed, to achieve them you must do one thing before the other: teleological explanations still rely on the natural ordering of events/actions in time. There is also ample psychological and linguistic evidence that unless explicitly marked to the contrary, either by tense or modality, or countermanded by pragmatic world knowledge, the antecedent-consequent order in a conditional encodes the natural casual/action/temporal or predictive order (Akatsuka, 1986; McCawley, 1980; Evans & Newstead, 1977; Evans, 1977; Evans, 1982). So unless there are contrary indications, it can be assumed that conditionals reflect this due to the natural ordering of constraints. However, the particular eductions, transitions between mental states, which constraints license can go in either direction. Henceforth, "Reasoning in the explanatory direction", will be used to refer to the use of a constraint to educt to a prior cause, reason, or *pre*condition.

The asymmetry between a constraint and its converse explanatory constraint is again given by pragmatic world knowledge. Furthermore, this knowledge influences the way a non-taxonomic constraint and its corresponding explanatory constraint can be used. Two issues are raised in reasoning in the explanatory direction. First, given (4.12) if the light is *not* on this is not necessarily explained by switch A not being turned, the electricity may be shut off, the bulb has blown etc. So the context sensitivity of the constraint invalidates inferences by *modus tollens*. Second, if the light is on this is not invariably explained by switch A having been turned. As it so happens, in my hall there is another light switch, lets call it B, which also operates the hall light. So, there is an additional constraint in operation, which acts as an additional possible cause of the light being on. If the light is on, it could have been because switch B was turned. In general, events tend to have many possible causes which legislates against unrestricted reasoning in the explanatory direction (cf. 3.4.1). Unrestricted explanatory reasoning from the non-occurrence of the indicating type (known) to the non-occurrence of the indicated type (unknown) or from the occurrence of the indicated type (known) to the occurrence of the indicating type (unknown), would only be rational in very constrained circumstances. This contrasts with the taxonomic constraint in (4.11), where the scarcity of albino ravens renders a *modus tollens* inference rational, but the knowledge of the respective domains renders an inference corresponding to affirming the consequent unreasonable.

Situation theory facilitates the regimentation of the distinction between taxonomic and non-taxonomic constraints which permits the ready classification of the asymmetry in the kinds of inferences warranted by each. Formal attempts to define a confirmation relation fail due to the incursion of content into the inductive process. Possessing a conception of laws as the structural relations in our world which license information gain, allows encoding these

contentful relations and provide some concept of how they mediate inductive inference (cf. the predictive cycle). In the next section, constraints, taxonomic and non-taxonomic, will be shown to provide a rational basis for some of the behaviour observed in Wason's task.

4.2.2 Disjoint vs unified rules and belief bias

In this section some empirical evidence will be considered for the distinction between the way subjects perform on the task modulo the very concept of a constraint and the distinction between taxonomic and non-taxonomic constraints. Prior to a discussion of the relevant experiments (Wason & Green, 1984), the standard selection task will be re-introduced.

Wason's (1966) task concerns how people assess evidence relevant to the truth or falsity of a rule expressed by means of a conditional sentence, normally using the *if...then* linguistic construction. Subjects were presented with four cards each having a number on one side and a letter on the other. On being presented with a rule such as, "if there is a vowel on one side, then there is an even number on the other" and four cards, one with each of the following letter showing on the upwardly turned face:

A(p) K(¬p) 2(q) 7(¬q)

they would have to select those cards they *must* turn over to determine whether the rule was true or false. The original assumption was that this assessment would be dependent on the form of the construction used to express the rule. Subjects responses should be predictable from truth tables which supply the meanings of the various logical terms. A conditional, $p \rightarrow q$ is true just in case either p is false or q is true, conversely it is false just in case p is true and q is false. So, the rule is true if and only if ("iff") *all* cards either have consonants on one side or *all* have an even number on one side, or some mixture of both such that each card has one or the other or both. Conversely, it is false iff *one* card has a vowel on one side and an odd number on the other side. Either way, only the "A" card and the "7" card must be turned over. If a subject is exhaustively trying to make the rule true then these are the only undecided cases. If a subject is trying to make the rule false then these are the only cases with the potential to do so. Typical results on this task were p and q cards (42%), p card only (33%), p, q, $\neg q$ cards (7%), p and $\neg q$ (4%) (Johnson-Laird & Wason, 1970). The preponderance of p, q and p only cards was taken as evidence for subjects adopting a verification principle, ie. they are looking just for pq instances.

Wason & Green (1984) identified a distinction directly related to that between taxonomic and non-taxonomic constraints as an important determinant of subjects performance in a reduced array version (RAST) of the selection task (Johnson-Laird & Wason, 1970). In the RAST subjects are only allowed to make selections concerning the consequent cards. In their experiment 3, the materials used were either coloured shapes and a rule such as:

(4.13) All the triangles are red.

which they referred to as the *unified* condition, or cards divided in two with a shape on one half and coloured on the other half, with a rule such as:

(4.14) All the cards which have a triangle on one half are red on the other half.

which they referred to as the disjoint condition. A "mental" version of the RAST was used. Subjects had to imagine that the shapes were in a case. Their task was to prove the relevant claim true by requesting information about the various coloured shapes using the smallest number of instances. Significantly more falsificatory responses were observed for the unified condition over the disjoint condition. On the suggestion of Johnson-Laird, Wason & Green also repeated the experiment correcting for sentence complexity (4.14 is considerably more linguistically complex than 4.13); the same pattern of results was obtained.

Rule (4.13) is directly analogous to (4.10), so analogous predictions would be expected of this taxonomic constraint. By parity of reasoning, with only consequent cards to choose from, subjects should select ¬red cards. An instance which is ¬red can not be a proper anchor for a triangle given the restriction in the indicated type, ie. they stand in a class inclusion relation, with the restriction as the superordinate category. The domain of red objects can be assumed to be larger by parity of reasoning with ravens example and so the red card is not turned. However, there is a distinction between (4.13) and (4.10): the objects in Wason & Green's experiment form a *closed finite* domain The domain is made explicit by the prepositional phrase in (4.15):

(4.15) All the red things *in the case*, are triangles.

In subsequent sections this will be identified as a further major determinant of subjects apparently falsificatory behaviour on this version of the task.

(4.14) is pragmatically *ambiguous*. The disjoint nature of the materials means that the rule is expressed as if it were a non-taxonomic constraint relating discrete events. The natural subject/predicate ordering has been broken up in the process. Both predicates now qualify different *parts* of a single object. This creates ambiguity because the only relation which

can exist between antecedent and consequent is class inclusion. That is, those cards with a triangle on one half are included in the class of cards which are red on one half. However, the disjoint expression of this taxonomic constraint appears to override this interpretation. And in the absence of any other constraints which could possibly block a prediction (explanatory direction) from *red* to *triangle* they request information relevant to confirming the rule.

How would (4.14) be regimented in situation theory? Despite its disjoint expression, the antecedent and consequent of (4.14) are about the same thing, ie. the cards. This satisfies the criteria for a taxonomic constraint. In chapter 2, taxonomic constraints were defined as those were the paramters of the indicated type were included in the indicating type. Argument roles and restrictions encode the *type-token* hierarchies in which the objects that can be assigned to these roles are located. Similarly, the minor constituents of a relation are *parts* of a state of affairs of a type identified by the major constituent. The relation between major and minor constituents encodes certain restricted *part-whole* hierarchies. Given a particular relation the minor constituents are the *logical* (or perhaps *analytic* would be more felicitous) constituents of the relation. It seems unobjectionable to use the same mechanism to encode *contingent* physical part-whole relationships. Allowing the card halves to function as argument roles (4.14) could be regimented as in (4.16):

T_Σ: *<<Card, one-half:* **x**, *other-half:* **y**; *1>>* & *<<Triangle,* **x**; *1>>*

T'_Σ: *<<Red,* **y**; *1>>*

(4.16) $<<[n_t]\Rightarrow, T_\Sigma, T'_\Sigma; 1>>$

(4.16) is still a taxonomic constraint. However, to grasp the inclusion relation subjects must treat the card as providing the unifying object of which the sides are parts. The disjoint expression of the rule seems to legislate against such an interpretation, which would be equivalent to suppressing the first conjunct in the indicating type, leaving just the restriction in the second conjunct. This of course, results in a non-taxonomic interpretation. Since, if the first conjunct is suppressed then it is no longer the case that the parameters of the indicated type are included in the indicating type. This is not to suggest a psychological mechanism but merely to illustrate the plausible interpretational effect of the disjoint rule expression.

Further empirical results seem rational form the perspective of the competence model. This is the work of Pollard (1979) and Pollard & Evans (1981) on *truth status* effects and *belief bias*. The asymmetry observed between:

(4.9) All black things are ravens, and
(4.10) All ravens are black

is exemplary of belief bias. (4.9) is not a hypothesis subjects are likely to believe to be true on the basis of their beliefs about the world. Rules believed false would be more likely to yield falsificatory responses. If you don't believe (4.9) you are likely to believe that there are black non-ravens and so look for non-ravens which are black. Pollard (1979) explicitly varied subjects familiarity with a pack of cards such that it was invariably the case that, say an A always had a 2 on the other side. Affirmative rules which suggested a correlation between A and cards other than 2, yielded significantly more falsificatory responses than rules which conformed with subjects experience of the pack. Pollard & Evans (1981) conducted a similar experiment using contentful or thematic material and observed similar results. It would appear that prior beliefs concerning whether a positive or negative correlation holds between antecedent and consequent will effect subjects performance.

Pollard and Evans (1981) suggest an *associational* explanation for these results. Learned associations between antecedent and consequent mediate subjects strategies for confirming rules. Being attuned to a constraint *just is* using the learned associations between events in the world to make predictions. Constraints may be positive or negative. In the later case a *preclusion* relation holds between these events (cf. chapt 2, chapt 6). It would be fair to say that what this thesis is attempting to do is work out a pragmatic (Evans, 1982:212) associational account of this data which is as rigorous as possible while allowing that (i) the standard task is inductive, (ii) task behaviour is reliant on pragmatic world knowledge, and (iii) the real world relations subjects' reasoning depends upon are context sensitive.

The distinction between taxonomic and non-taxonomic constraints and their usual modes of expression appear to critically influence subjects selection task performance. When the rule is expressed in a way which permits conceptualisation of the described situation in terms of instances, the class inclusion relation appropriate to a taxonomic constraint is evoked. Inferences are therefore licensed which apparently concur with falsification. However, when the rule is expressed disjointly, thereby identifying a non-taxonomic constraint different modes of thought are engaged. Rather than identifying instances which have the various properties, this elicits a predict and explain strategy appropriate to determining whether one discrete event is a good predictor of another (cf. above). In all standard selection tasks the requirement to display only one side of the card or otherwise restrict information, means that the rules are always expressed disjointly. This ambiguity will prove central to explaining the results on standard versions of the task. The closed domain employed in Wason & Greens's

task will also be identified as influential in determining task performance. This will be discussed in the next section, where it will be shown that certain paradoxes which beset confirmation theory (cf. chapter 3) may appear less paradoxical in small surveyable domains. It will be shown how the factors involved provide a rational basis for subjects behaviour in versions of the task similar to Wason & Green's (1984). The purpose of this discussion will be to highlight the differences in these tasks and the standard version. The remainder of the thesis will then concentrate solely on explicating the rational basis of subjects divergent behaviour in standard versions of the task.

4.3 Confirmation, surveyable domains and the COST

The modal response of *p* and *q* cards observed in standard versions of the selection task was taken to argue that subjects are attempting to *verify* the rule (Wason, 1966; Wason & Johnson-Laird, 1972; Evans, 1982). The normative competence model assumed in interpreting this data has been Popper's principle of falsification. This lead to a verificatory procedure to be labelled a *bias*. However, attempts to formally define a confirmation relation were discussed in the last chapter. The reasons for the failure of these formal systems are highly instructive concerning the kinds of strategies appropriate to *closed domain selection tasks* like Wason & Green's (1984).

Nicod's criterion involved a property restriction on the domain of the implicit universal quantifier. Another type of restriction is an *objectual* restriction, where the domain of the quantifier is restricted to a limited finite set of objects. Von Wright (1957) introduced such a restriction, known as the *Postulate of Completely Known Instances*, in articulating the elliptically present premises required to turn induction into deduction. Goodman also makes use of the distinction in explicating the difference between an *accidental generalisation* and a scientific law. In order to confirm an accidental generalisation, eg. all the coins in my pocket today are silver, all the coins in my pocket would have to be surveyed before it could be confirmed. However, a putative scientific law may be accepted without all the evidence being available.

The domain of the quantifier in Wason's task is explicitly restricted to just the four cards. Therefore, the rule would appear to be a prototypical accidental generalisation. However, the cards are not completely known so without turning the cards the truth of the rule cannot be deduced. Nonetheless, the restricted domain does appear to lessen the paradoxical status of the Ravens example. In a domain of only 4 cards, relative to the rule *If vowel one side,*

even number on the other, the information that a card which has not got an even number on one side also has not got a vowel on the other, is confirmatory. It does not disconfirm and hence confirms. What it is rational to treat as confirmatory in a finite surveyable domain may be irrational in a potentially infinite or at least very large and unsurveyable domain. In the latter case the set of potential non-disconfirming instances will be prohibitively large. For example, the set of non-black non-ravens is far larger than the still unsurveyable set of ravens. The paradoxical status of the ravens example may only be attributable to the unsurveyable domains of a scientific law. Despite this subjects do not look at the cards which could correspond to $\neg p$, $\neg q$ instances (K & 7), rather they appear to verify, ie. look for only potential p, q instances.

However, it has been observed (Beattie & Baron, 1988) that selection of an A-card on the task is ambiguous between verification and falsification. This card could be selected because of its potential to *falsify*. It can also be observed that in tasks like Wason's & Green's (1984, expt. 3/4), where subjects appeared to falsify and there was a small finite domain of objects, selection of $\neg q$ instances is similarly ambiguous. Subjects could be asking for this instance because if the rule is true they expect $\neg p$ on the other side, ie. they are looking for $\neg p$, $\neg q$ instances to *confirm*. Tasks like the selection task and Wason and Green's (1984) RAST do not permit identification of whether subjects are looking for $\neg p$, $\neg q$ instances to confirm or p, $\neg q$ instances to falsify. Hence, Wason's & Green's result may be due to the finite domain of cards which encourages a confirmation strategy (as opposed to verification). The strategies subjects' employ can not be distinguished between on the basis of this data.

One way around this problem is to elicit protocols concerning subjects justification for selecting a card. However, it has often been observed that subjects' verbal justifications do not match there actual task behaviour (Evans & Wason, 1976; Evans, 1982). This has lead to the suggestion of dual processes in subjects reasoning (Wason & Evans, 1975; Evans, 1980a, 1980b). Rapid automatic processes operate to determine subjects initial responses but higher level conscious processes are implicated in their subsequent verbal justifications. This distinction will have a role to play later on in discussing the processes which are responsible for determining subjects rule interpretations. However, Wason's and Evans' arguments concerning protocol data tend to argue against its employment in determining the strategies subjects are actually adopting during the selection phase of these experiments.

However, Beattie & Baron (1988) introduce a procedure which could potentially distinguish between confirmation in small finite domains and falsification. They employed a similar

task to Wason & Green (1984) (experiment 3). In both tasks the domain is explicitly closed, ie. objectually restricted. To mark this I propose a new term for these experiments: Closed domain Selection Tasks (COSTs). Wason & Green's version employed a reduced array: $COST_{ra}$. Beattie & Baron's instructions were to fully identify the cards which they thought would be informative, specifying both sides of the cards. They call this procedure a multi-card task: $COST_{mu}$. Four cards were drawn from a pack of 21 cards perming all the possibilities of numbers 1 to 3, and letters A to G. These four cards were placed in a bag. Subjects had to construct, write down, all the cards they would want to see that could be informative concerning the rules truth or falsity. Given an affirmative rule, *if A, then 2*, if they were falsifying, they should look for $p, \neg q$ instances, and if they were confirming in a finite domain they should look for $\neg p, \neg q$ instances. Subjects predominantly chose falsifying instances, almost no $\neg, \neg q$ instances were chosen. However, the *same* subjects invariably made the standard verification selections in an abstract selection task.

In spite of this apparently conclusive data, perhaps what subjects are doing is not *purely* falsificatory. Beatie & Baron classified subjects responses as falsificatory on the basis of the following observation:

> (4.17) "Many subjects argued that one should first look for any potentially falsifying cards, indicating that their presence would prove the rule false. However, if none was found, one should then check that at least one A-card was present, lest the rule be vacuous. We felt that this argument successfully demonstrated the principles of falsification and hence scored it as correct"

However, the strategy subjects reveal here is wholly consistent with confirming in a finite domain. Subjects could be asking for instances with the potential to disconfirm, as long they don't, *they confirm*. But there must be at least one positive instance "lest the rule be vacuous".

In real world inquiry in a finite domain this strategy is wholly rational. For example, in a hardware shop Johnny is asked to check that all the 1" pipes are threaded. Johnny reasons thus: he knows that most of the pipes are threaded (they are supposed to arrive from the factory that way), given this information he is far better advised to check all unthreaded pipes to check they are not 1", than all 1" pipes to check they are threaded. As long as this is the case, he can report back that indeed all the 1" pipes are threaded. But if the reason for the check is that threaded 1" pipes are needed, then he better make sure there are at least some 1" pipes. In a finite domain this procedure leads to *certain* knowledge of the *truth* of the hypothesis. It also relies on some prior knowledge of the likely distribution of threaded and unthreaded pipes. But this is incidental when it is considered that the accepted

grounds for adopting a falsificatory strategy is that *one can never be certain that a hypothesis is true, but one can falsify with certainty* (Popper, 1959). *But* in a small surveyable domain this is false. Falsification is considered rational *because* the domains under consideration are unsurveyable and hence truth can never be guaranteed. But in a small finite domain, relative to knowledge of prior distributions, *confirming* by discounting potentially falsifying instances is rational.

Wason's task was devised as a test of subjects ability to evaluate scientific hypotheses (Wason & Johnson-Laird, 1972). Such hypotheses are stated *without* objectual restrictions on the domain of the implicit universal quantifier. Scientific laws and constraints are required to allow prediction in as yet unknown, unsurveyed domains and not just particular restricted, surveyable domains. The strategies appropriate to each are different. The putative evidence for falsification functions equally as evidence for a particular *confirmation* strategy. To demonstrate falsification would require subjects not to accept any positive instance as confirmatory. This renders falsification implausible as a psychological procedure, albeit arguably the counterintuitive procedure scientists should adopt with regard to their hypotheses, such an unremitting commitment to fallibilism is not wholly rational. The context sensitivity of most constraints or laws means they may not be susceptible to direct falsification (cf. below and Putnam, 1974).

Evidence from the COSTs is equivocal with regard to whether subjects are adopting a falsificatory procedure. If the domain is closed then strategies of confirmation may be elicited which appear consistent with falsification. It is important to identify the factors which are responsible for the divergent behaviour on the standard selection task. The rest of this section is devoted to determining the relevant respects of similarity and difference between these tasks and the standard version. It will be argued that the circumstantial manipulations of the task affect subjects interpretations of the rules, and therefore their subsequent strategies of confirmation.

In Beattie & Baron's task subjects knew all the relevant instances and had to *evaluate* whether an instance would be informative with regard to the rules truth or falsity. In Wason & Green's task subjects could ask about only the colour of the card. Wason & Green (1984) also observed facilitation on the task when using a more concrete procedure where information about the cards was restricted by the use of masks, which argues strongly that the unified rule condition was responsible. However, Beattie and Baron used only *disjoint* rules but still observed a facilitation in a COST. Which argues that the closed domain is also responsible for the adoption of apparently falsificatory procedures. So, there

was facilitation for disjoint rules in a COST in Beattie and Baron (1988), *but* a corresponding failure to observe facilitation in a COST using disjoint rules by Wason & Green (1984). The discrepancy can be plausibly resolved by considering the response procedures employed. Beattie & Baron's task was not a RAST, subjects had to *evaluate* pairs in imagination as relevant or not to the truth or falsity of the rule. The closed domain functions to elicit the consideration of "falsifying" pairs. In Wason & Greeen (1984), subjects had to identify one half of the pair, which is equivalent to having to use the rule to *explain* whats on the other half.

Beattie and Baron (1988) observed that a frequent misunderstanding made by subjects was the belief that the four cards in a selection task represented a sample of a much larger population. They discuss this factor as a possible determinant of the discrepancy observed between the same subjects $COST_{mu}$ performance and standard selection task performance. However, they conclude that it is probably not the crucial factor. They reason that in the four card problem there are many more potential combinations as the domain of numbers and letters is not fixed as in the $COST_{mu}$. On the back of any letter there is a potentially infinite number of possibilities and 26 on the back of each number. They argue against the hypothesis that respective domain size is important based on an unpublished experiment where they restricted the letters to A & B, and the numbers to 2 & 3. They observed worse (less "falsification") performance, albeit not significantly.

This procedure fails to identify the relevant dimensions of the domain. It is not the number of possible numbers or letters which is important, but the size of the domain of possible instances functioning as *tokens* of the *type* mentioned in the rule. The rule is inherently general, it states that cards which are tokens of the type such that there is an A on one side, are also tokens of the type which has a 2 on the other side. It is the size of the abstract classes, relative to the fact that there may be many tokens of the stated type, which is important not the fact that many numbers are members of the class not-2. The cards are drawn from a pack containing the set of possible instances which represents the appropriate dimension of generality that needs to be controlled. The critical manipulation would be to remove the generality and observe more "falsificatory" responding. This could be achieved using the past tense indicative, rather than the present tense. Conditionals in the present tense and the indicative mood (all rules used in this task are of this linguistic form) usually express generalities. However, take:

(4.18) If there was an A on one side, then there was a 2 on the other.

The task-cards are then hypothetical instances of this single card, subjects have to identify

which hypothetical instances of this single card could prove the asserter of (4.18) right or wrong. (This suggestion is not actively explored here, but subsequent work will look at this hypothesis).

It would appear that in the COST the closed domain functions to license a strategy of confirmation appropriate to a limited surveyable domain which allows instances which do not disconfirm to confirm. However, a unified rule expression also permits the identification of a falsificatory strategy even if it is ambiguous as to whether the domain is open or closed. In most versions of the selection task the rule is disjointly expressed and on the assumption that subjects take the rule to apply to an open domain then the predictive cycle strategy is appropriate. This will form the basis of providing a rational foundation for subjects standard selection task behaviour in the next chapter. The task versions surveyed here indicate that various manipulations are affecting subjects actual interpretations of the rules, including the domain of the implicit universal quantifier. However, these manipulations are not all concerned directly with the rule expression it self. It seems as though *circumstantial* pragmatic features of the task are entering into subjects rule interpretations which subsequently affect their strategies of inquiry. This argues that logical semantics may provide to narrow a definition of "interpretation" to function as a standard for judging psychological performance. It would appear that such pragmatic factors are directly implicated in subjects rule interpretations. In the area of sentence processing it has been established that any distinction between semantics and pragmatics is purely a matter of formal convenience and need not imply autonomy at the cognitive level (Crain & Steedman, 1985; Marslen-Wilson & Tyler, 1980). Perhaps then *circumstantial* features of the specific tasks enter into subjects rule interpretations. Such circumstantial determinants of subjects performance may also derive from their prior beliefs concerning sizes of respective domains, and the *relationships* asserted to exist between antecedent and consequent.

In Barwise & Perry's terminology interpretation also depends upon an individual interpreter's *resource situation*, ie. the disposition of the environment (the task requirements) and his prior beliefs. The strategies appropriate to fixating conditional beliefs will vary as a function of these manipulations because they affect the interpretation of the rule. This implicitly rejects the idea that there is one and only one normative strategy for fixating conditional beliefs. However, the bulk of the variation may well operate in small finite domains relative to differences in known *a priori* distributions. For example, Johnny may have no prior knowledge concerning whether there are more threaded or unthreaded pipes. Concern then centres on the respective distributions of 1" and other pipes. If there are likely to be far less 1" pipes than others, then looking for 1" pipes to check that they are

threaded is more rational than checking the unthreaded pipes to see that they are not 1" (cf. Klayman & Ha, 1987, and below). However, in open domains where often such *a priori* knowledge is unavailable fewer options present themselves.

By considering the paradoxes which emerge for confirmation theory it has been shown how an account of task performance in the COST can be derived which does not indicate that the manipulations performed involve the facilitation of purely falsificatory behaviour. Closed domain tasks are however, non-standard. In the tasks standard form a disjoint rule is employed which countermands the consideration of instances in favour of a relation between discrete occurrences. Only the standard selection task will be considered in the rest of this thesis.

In the discussion of confirmation theory (cf. chapter 3), it was mentioned that Belnap had defined a connective he called "conditional assertion". In the psychological literature this concept has been appealed to in accounting for the observation of a defective truth table (Wason, 1966). In some tasks, subjects behaviour indicates that they believe false antecedent instances of a rule are irrelevant to its truth or falsity. *Prima facie* partial interpretation provides a rational basis for this behaviour, which also implicates this observation as the principle determinant of the why falsificatory behaviour is not usually rational. However, in the psychological literature the observation of defective truth tables has not been influential. The reasons for this will be explored in the next section where it will be shown that partial interpretation has far reaching consequences for the rational interpretation of this data.

4.4. Partial interpretation and defective truth tables

If Johnny finishes his check and reports back to his boss that indeed all the 1" pipes are threaded *but* he didn't check to see whether there were any 1" pipes, what would his bosses reaction be on arriving at a job and discovering he had no 1" pipes? I suggest his response would be unprintable. If the rule is vacuous, the boss is unlikely to concede its truth. This is no doubt mediated by the fact that a rule which is "true" vacuously does not permit the effective planning of subsequent action. Or, perhaps, more felicitously it may elicit inappropriate action. If Johnny's boss knew the rule was "true" but vacuous, rather than go on the job, he would have ordered more 1" pipes. Possessing the right information is crucial to planning appropriate action. Since this is the case, just looking for $\neg p$, $\neg q$ instances may well be confirmatory, but only if there are some p instances.

Wason (1966) proposed that subjects may reason with a defective truth table which assigns the value irrelevant ("?") to false antecedent instances. Several *truth table* tasks have subsequently borne out this observation. In a truth table task subjects are either asked to *evaluate* instances as true, false or irrelevant (Johnson-Laird & Tagart, 1969; Evans, 1975; Evans & Newstead, 1977; Evans, 1983b), or are asked to *construct* verifying or falsifying instances (the irrelevant category being inferred, cf. chapter 6) (Evans, 1972). Wason's conception of a defective truth table was based on the Quine-Rhinelander idea concerning conditional assertion which motivated Belnap's (1970) definition of a non-contraposing conditional connective. This proposal may get round Johnny's problem. $\neg p$, $\neg q$ instances are irrelevant, and hence do not disconfirm. Perhaps then it can be allowed that if he only discovers $\neg p$, $\neg q$ instances he can report that the rule is confirmed but vacuously so, and therefore the truth of the rule is undecided. This proposal is legitimate since the confirmation relation relates an instance to a generality, it does not relate *all* instances to that generality. If the domain extends to all pipes, wherever they may be then a different question is being raised which goes beyond Johnny's epistemic resources to decide upon. Hence, although each instance confirmed, taken together all the instances (in the pipe box) do not legislate positively with regard to the truth of the hypothesis.

Situation theory captures this phenomenon by the use of partial interpretation relative to a limited situation. The rule Johnny is checking concerns a taxonomic constraint about 1" gas pipes which should be threaded.

T_Σ: *<<1"_pipe, x; 1>>*

T'_Σ: *<<Threaded, x; 1>>*

(4.19) C: $\langle\langle [n_t] \Rightarrow, T_\Sigma, T'_\Sigma; 1 \rangle\rangle$

Let the situation under consideration s be that given by the content of the pipe box. If $s \models C$, then for any anchor f such that $s \models T[f]$ it is also the case that $s \models T'[f]$. Whether a constraint holds in not reducible to the right hand side of this statement, necessary and sufficient conditions are not being specified. First, there may be no anchor for the indicating type, in which case C may hold or not, but equally there may be an unthreaded 1" pipe but other relevant conditions are operative which means that C still applies, ie. there could be a particular background type was not fixed in s. For example, suppose Johnny, observes some 1" pipes which are unthreaded. Relative to the situation s as he currently individuates it, $s \models C$ is false. But he did notice that all the unthreaded 1" pipes were bends. So when he reports back to the boss he says, "Not all the 1" pipes were threaded, but all the unthreaded pipes were bends", the boss replies "Thats OK, the bends are never threaded".

This new information changes the situation, as long as the unthreaded 1" pipes are bends, the new situation s' supports C. Partial interpretation captures the context sensitivity of peoples reasoning.

Wason's proposal that people possess a defective truth table has not been influential in psychological theorising about the selection task, despite the existence of corroboratory evidence. This appears to be due to the assessment that it should not affect subjects performance. This clearly contradicts the view presented here, hence an argument is required to demonstrate why this assessment may be wrong.

The above interpretation relies on capturing the appearance of truth value gaps in a particular way. Certain partial states of the world enter into the semantic analysis to encode context (as observed in chapter 2, Veltman (1985, 1986) achieves a similar result by appealing to epistemic information states). Relative to these bits of the world the truth functions which take propositions into truth values are partial. However, some logicians have suggested that a third (or more) truth values should be admitted in order to capture truth value gaps. Rarely do the other value(s) share the same status as *true* and *false*, they tend to have rather specific connotations dependent on the philosophical/logical purpose to which they are being put (cf. Haack, 1975:Appendix). For example, Kleene's (1952) third truth value has a meaning given by the concept of a partial function, it simply means *undefined*. This is because he was proffering an analysis of recursive functions. Reichenbach's (1944) third truth value was assigned to sentences "about entities which, in certain conditions, it is impossible to measure" (Haack, 1975:172). This was because he was concerned with issues surrounding the measurement problem in Quantum physics. Truth tables for *all* the connectives are defined in these systems.

Johnson-Laird & Tagart (1969) observed that subjects responses in a truth table task were consistent with a defective truth table for the conditional. They argue that on this interpretation the contrapositive does not hold but that the $\neg q$ card is, "a required choice on any reasonable interpretation of the conditional, including even the non-standard interpretation of the present experiment." (Johnson-Laird & Tagart, 1969:372), ie. falsificatory behaviour is still to be expected. Let us assess this argument. The contrapositive fails immediately: for the equivalence to hold the formula $p \rightarrow q$ and $\neg q \rightarrow \neg p$ would have to be assigned the same truth values for every combination of assignments to p and q, but given the defective truth table this is no longer the case. However, as usually articulated falsification relies on the validity of modus tollens, ie. $p \rightarrow q, \neg q \models \neg p$. *Prima facie* this is no longer a tautology because there is *no* assignment which satisfies both $p \rightarrow q$ and $\neg q$, they can never be

true together. p -> q is only true (on the defective truth table account) when q is true (hence $\neg q$ false) and p true (hence $\neg p$ false). Unless the definition of tautological implication (|=) (cf. Enderton, 1972) is altered, this means that modus tollens is not valid when a defective truth table is countenanced. However, to make a proper determination of this claim requires the irrelevant category to ramify throughout the logic in the appropriate way. It may be the case that when the connectives are fully defined for a third truth value then, modus tollens is valid. For example, in Kleene's strong 3-valued system modus tollens is valid; in Reichenbach's scheme for *alternative implication* it is not (cf. Haack, 1975). In Veltman's (1985, 1986) data-logic, modus tollens is generally invalid, except in the limiting case when a chain of information states is *complete* and "->" gets the meaning of material implication.

Does falsification rely on modus tollens? As usually articulated, yes. But falsification concerns hypotheses which are universally quantified, as are all the rules in the selection task. As long as the same instances of a conditional are false, then a universal claim stated over some set of objects will be false dependent on the semantics of the quantifier. Relative to the domain of the model, a universal claim is true if all substitutions of the objects for the variable bound by the quantifier possess the property made in the claim. So, in the domain of ravens, a claim $\forall$xblack(x) is true if each raven is black, it is false if one raven is not black. Restricted quantification as articulated in Nicod's criterion is equivalent to the claim that the antecedent implicitly identifies the *non-empty* domain of the appropriate model. But what value is assigned to the quantifier when the domain is *empty*. This is just the same question as what value do we assign the conditional when the antecedent is false!. By convention and to preserve bivalence, in these vacuous circumstances the quantifier is assigned true. But this assignment is unlikely to impress Johnny's boss, if there are no 1" pipes he needs to know that the rule is vacuous.

The interpretation of the quantifier depends on the domain of the model. When this is fixed, the appropriate *context* of the generalisation has been identified and when this is the case a $p, \neg q$ instance falsifies. Let us return to Johnny, who has just observed that some 1" pipes are unthreaded *but* that they are all bends. Relative to the domain of all 1" pipes in the box, the claim that "all 1" pipes are threaded" is false, but when his boss replies, "Thats OK, the bends are never threaded", the domain over which the generalisation is stated has implicitly changed to all pipes in the box less the bends, in which case "all the 1" pipes are threaded" is now true. Normal human reasoning is *non-monotonic*; additional information can render sound formerly unsound inferences or vice versa. The logical axioms and inference rules of a system with quantifiers are stated over *all* models. Non-logical axioms are

going to be tied to particular interpretations, ie. models. Change the model and an inference licensed by one non-logical axiom will no longer be valid. Systems like situation theory and Veltman's data-logic attempt to capture this behaviour via partial interpretation, partial bits of the world or knowledge states explicitly identify the relevant context of interpretation.

In general, falsification is only guaranteed to arrive at certain knowledge once the domain of the generalisation has been fixed *in advance.* This is not an *objectual restriction* but a *property restriction.* Falsification is an open system strategy, *but* it assumes that all the relevant properties of the objects involved have been fixed. So, in the case of Johnny the extra property of being bends is not admitted as relevant, so the generalisation is just false. But in real human inquiry, people can never be sure *a priori* what the relevant properties are. Rather, as in the predictive cycle, instead of taking the hypothesis to be falsified, the predictive failure triggers further inquiry into the proper domain of the generalisation (cf. Johnny noticed that it was particularly salient that all the unthreaded 1" pipes were bends). With regard to the *certainty* which can be obtained, this places falsification on the same level as confirmation. Inquirers can be no more certain that a hypothesis is false, given a predictive failure than they can be certain it is true given a predictive success. However, given a predictive failure the process of delimiting the domain of applicability can be initiated. If, in the long run, all such attempts fail, then the hypothesis will (i) in the case of individuals, simply fall into misuse, or (ii) in the case, say of the scientific community, become a Kuhnian *puzzle.*

The possible consequences of partial interpretation as evidenced in defective truth tables apply to an argument put forward by Johnson-Laird and re-iterated by Evans (1972). The argument is that even if modus tollens is no longer valid, as the result of a defective truth table, inferences corresponding to modus tollens are still valid because *reductio ad absurdum* can be used to derive similar inferences. Two observations need to be made. First, the observation of a defective truth table can not be allowed just to affect the conditional at pain of sacrificing compositionality. Once a third truth value or partial interpretation has been conceded it must be allowed to ramify through out the logical system. Second, *reductio ad absurdum* depends on another logical law: the *law of the excluded middle.* However, in a three valued system the law of the excluded middle is not normally valid. For example, it is not an axiom of Kleene's strong three valued system (Haack, 1975), although modus tollens holds in that system. Moreover, it was a rejection of the law of the excluded middle and its role in apparently vacuous mathematical existence proofs which motivated the development of intuitionistic logic (Haack, 1975). Similarly, in Veltmans (1985) data logic

the law of the excluded middle only holds when a chain of information states is complete, ie. all the information relevant to the truth or falsity of a claim is available and all relevant propositions can be assigned a determinate truth value.

The assessment of the psychological implications of the observation of "defective" truth tables for subjects reasoning may have been premature. The logical consequence of this observation is a view of the interpretative process which is inherently partial and which fails to license falsification as any more rational than confirmation. Moreover, the consequences are consistent with intuition (cf. Johnny and the pipes), the model of the predictive cycle, and with the psychological data so far reviewed. The importance of this observation should not go under-estimated. In the chapter 6, data on Evan's negations paradigm will be introduced. Evan's (1972, 1982, 1983b) provides arguments for retaining standard logic in the interpretative component of the cognitive mechanism, on the grounds that the observation of irrelevant responses can be put down to a non-logical response bias, called *matching*. However, once the rational basis for this behaviour has been identified, via the specification of an appropriate competence model, the data must be re-interpreted in that light. This will be done in the following chapters.

4.4.3 Philosophical postscript on falsification

The observations on the inadequacy of falsification have been principally motivated by semantic considerations concerning context dependence and partial interpretation. In this section therefore, it will be shown how these semantic considerations directly reflect long standing doubts within the philosophy of science concerning the coherence of falsificatory procedures. It will also be shown how this affects some recent psychological arguments concerning the *positive test heuristic* (Klayman & Ha, 1987).

A paper by Hilary Putnam (1974) best exemplifies the problems which beset falsificationism. The paper demonstrates that Popper is wrong to conclude that science can do without induction. Putnam addresses the question of how predictions are derived from theories? Using Newton's Universal Theory of Gravitation (*UG*) as an example Putnam demonstrates that the derivation is not direct, ie. predictions are not deductive consequences of theories alone, as they were characterised above in discussing the work on confirmation theory. Rather predictions are only derivable from a theory taken in conjunction with a set of boundary or initial conditions, which Putnam calls *Auxiliary Statements* (*AS*). Typically these constitute certain simplifying assumptions, eg. if we take *UG* and the following *ASs*:

(i) No bodies exist except the sun and the earth.
(ii) The sun and the earth exist in a hard vacuum.
(iii) The sun and the earth are subject to no forces except mutually induced gravitational forces.

then from their conjunction Kepler's Laws can be deduced. Moreover, "By making (i), (ii) and (iii) more "realistic" - ie. incorporating further bodies in our model of the solar system - we can obtain better predictions" (Putnam, 1981:65). However, the *AS*s do not constitute part of the theory, they are simplifying assumptions about what the theory applies to, or its explanatory domain. The laws in question are the covering laws of a unified theory, which as Cartwright observes are unlikely to be true. Only the *ceteris paribus* laws which connect these covering laws to the actual phenomena are likely candidates for truth.

This effects Popper's doctrine immediately. Theories are *not* strongly falsifiable. The falsity of a prediction can not percolate directly back too the falsity of the theory: the *AS*s may be wrong. For example, *UG* was accepted as unquestionably true by the scientific community for over 200 years. Yet, when predictions concerning the orbit of Uranus were derived from *UG* in conjunction with the *AS* that all the (then known) planets were all there were, they turned out to be wrong. Rather than regard *UG* as therefore falsified, Leverrier in France and Adams in England predicted that there must be another planet. And indeed another planet, Neptune, was observed in 1846. So, in this instance, not only did scientists not falsify, they were right not to. It was similarly correct not to regard the perihelion of Mercury, another predictive failure, as falsifying the theory. As in the case of Uranus, the possibility could not be dismissed that there were other unknown forces acting on the planet. From examples like this, it is clear that what scientists do (their inferential practices), simply do not conform to the Popper's falsificationary prescriptions, but they are wholly consistent with the standard view "that scientists try to *confirm* theories and *AS*s by deriving predictions from them and verifying the predictions." (Putnam, 1981:68).

Putnam goes onto demonstrate the compatibility of these observations with the Kuhnian concept of a *paradigm*. A paradigmatic theory is one that is not allowed to be falsified. "Normal" science goes on within a paradigm by assuming the veridicality of the theory and trying to determine what the relevant *ASs* are, given a body of "to-be-explained" data. Ironically, despite Kuhn's emphasis on revolutionary science, ie. when a new theory comes to predominate, his views demonstrate what an essentially conservative enterprise science must be to attain progress.

Putnam's closes the paper with some observations on practice which can be interpreted as

highlighting the fact that a sound inference should enable successful practical action, ie. action in the real world. The generalisations from which inferences are made must accord with the world sufficiently well to facilitate action within it. The extent to which it does, increases confidence that they are sound. This can be the only real criteria of judging a theory. For Putnam and Goodman attempts to formalise this testing of theories, by eg. Carnap, Popper and Chomsky, are simply misguided. The relevant auxiliary assumptions in Putnam's terminology translate directly into the background conditions which need to be present for a constraint to hold in a situation. Once the appropriate eduction is performed and found to yield a predictive failure, this does not entail that the constraint does not hold, it could be that the relevant background conditions were not present. To constitute sound action guiding rules, the eductions performed by a cognitive agent must be grounded in the actual structure of the world, only then will efficient and effective action be possible.

These considerations from the philosophy of science also bear on some recent arguments put forward by Klayman and Ha (1987). They argue that positive tests, ie. looking for instances which verify, may indeed be rational *because* such a procedure can lead to a higher probability of receiving falsification. They point to an important distinction between modes of disconfirmation. An inquirer may either seek instances which are predicted not to have a certain property, or seek for instances which are most likely to falsify rather than verify the hypothesis. This point is demonstrated by appeal to Wason's (1960) 2-4-6 rule discovery task. Subjects are presented with a sequence of numbers and must formulate a hypothesis concerning the target rule known only to the experimenter which describes this number sequence. The target rule could be as general as "three consecutive even numbers" or just "ascending numbers". Subjects must then suggest possible sequences to determine whether their hypothesized rule matches the target rule. The experimenter provides feedback concerning whether or not the suggested sequence matches the target rule. Once subjects have generated a hypothesis, they typically ask for verificatory sequences rather than those which are false instances of their hypotheses.

Klayman & Ha demonstrate that since subjects are ignorant of the target rule they are also ignorant of the relationship between their hypothesized rule and the target. Four possibilities present themselves. Since each rule can be thought of as defining a space in the domain of possible triples the relationship can be described as holding between sets. The hypothesized rule is either included in the target set, or vice versa, they overlap, or they are totally disjoint. The desiderata is, of course, perfect coincidence. Given these possibilities, Klayman & Ha observe that in some cases it is possible to receive conclusive falsification using either positive or negative tests (disjoint or overlapping), but in one (hypothesized set

included in target set) only negative tests yield conclusive falsification, and in the remaining case (target set included in hypothesized set) only positive tests can yield conclusive falsification. They then demonstrate that under various boundary conditions, an assumption of a minority phenomenon (ie. the probability of the target rule applying in the population is less than 0.5) and an assumption that the probability of the hypothesis is roughly equal to the probability of the target, then it can be shown that the probability of obtaining falsification using positive tests is higher than using negative tests. The tests described all involve what Klayman & Ha (1987) refer to as Htests, ie. testing instances hypothesized to (not) possess the target property. However, in real world inquiry concern may equally focus on Ttests, ie. testing instances known to (not) possess the target property to determine whether they also possess the hypothesized property. By parity of reasoning with Htests they also show that +Tests are more likely to result in falsification than -Ttests.

There are several points to be raised concerning this demonstration of the falsificatory power of positive testing. First, the analysis assumes as a premise that falsification is the rational procedure subjects should adopt, be this achieved by negative or positive tests. However, the arguments from both partial interpretation and the philosophy of science invalidate this claim. Since this is the case the rest of their argument need not be taken as an adequate characterisation of the rational basis of subjects' inferential behaviour. Reiterating Popper (1959), they claim that:

> (4.20) "Put somewhat simplistically, a lifetime of verifications can be countered by a single conclusive falsification, so it makes sense for scientists to make the discovery of falsifications their primary goal." (Klayman and Ha, 1987:214)

However, the examples from the philosophy of science indicated above show this claim to be false. Within periods of normal science, which predominate, it is not rational to allow putative falsifications to penetrate to basic theory. The range of coverage possessed by the theory can render such an inference from a falsification in an isolated domain wholly irrational. The History of Science is replete with examples of putative falsifications being ignored, suppressed or covered by ad hoc hypotheses. The perihelion of Mercury was ignored for more than a century, the observation of double refraction was covered by the ad hoc addition of "sides" to the particles in Newton's corpuscular theory of light. The Michelson-Morley experiments were only taken as crucial some 30 years after the event. The list is endless. Only once a better theory is available would it be rational to abandon the old.

It was also observed above that Popper underestimated the force of evidence which normally accompanies the discovery of a rule/hypothesis. The processes by which people

discover likely hypotheses are many and varied but in the 2-4-6 task rule discovery is little more than guess work. Actual rule discovery in science is a far more sophisticated process which often imbues a hypothesis with some sound reasons to believe it. For example, Kepler's laws which describe the planetary motions was tested and matched against the wealth of data available from Tycho Brahe's observations. The form the laws could take was tightly constrained by the data, but it still required a great deal of intellectual effort to derive elliptical orbits. This contrasts radically with the 2-4-6 task in which the form of the rule is massively undetermined by the data. In deriving the Universal laws of gravitation, Newton already possessed Kepler's three laws to constrain the form they could take, but the inverse square law was still far from transparant. In deriving the equations which govern the behaviour of electromagnetic phenomena, Maxwell discovered an analogy between the laws he was be constrained to employ and the laws which describe the phenomena of light, thereby unifying two disparate domains of inquiry. In each case, (i) the rule discovered embodied its own evidential support via its ability to cover the phenomena, and (ii) the intellectual effort which went into their discovery was immense. Just coming up with one rule which achieves *some* coverage is a formidable achievement. Once a rule is discovered then further assessing its range of application is by far the most rational strategy to adopt. And indeed if a anomaly is detected, eg. the orbit of Uranus, it is best ignored until a better theory is forthcoming.

Klayman & Ha (1987) also apply the +test heuristic to the selection task. It is shown how +Htests and +Ttests could lead to the standard selections on the abstract version. However, they observe that because the rule only describes a sufficient condition there is only one falsificatory instance and therefore subjects should not really adopt the +test heuristic but use a -Ttest. They also propose that thematic content may facilitate this realisation. In general subjects abstract task behaviour is still deemed irrational, precisely because the assumption is still made that they *should* falsify. In the following chapters it will be shown how the predictive cycle and the concept of interpretation embodied in situation theory provides a unified rational basis for subjects behaviour on this and related tasks.

4.4.4 Syntactic vs semantic proof procedures

Johnson-Laird (1983, 1986a, 1986b) argues for an asymmetry between a connective's logical properties and its meaning. Syntactic rules of inference, eg. modus ponens and modus tollens, represent logical properties and truth tables, semantic properties. An asymmetry is observed between these properties insofar as, "a statement of the truth conditions of

conditionals constrains the form of inferences that are valid, but a statement of the form of valid inferences leaves conditionals open to a number of distinct semantic interpretations" (Johnson-Laird, 1986a:57). Given the truth table for the material conditional both modus ponens and modus tollens will be sound. But if these two rules of inference are taken to govern the conditional then it could have the truth conditions of either the material conditional or a defective truth table. It is conceded that on the latter interpretation contraposition would no longer hold but both rules of inference would remain sound.

Johnson-Laird & Tagart (1969) is cited as the relevant reference which establishes this claim. However, in that article no argument to this conclusion is offered. It was argued above that modus tollens is *not* sound under the defective truth table interpretation. To reiterate, propositionally for any inference to be valid the premises must tautologically imply (|=) the conclusion, ie. whenever the premises are true so is the conclusion. On the truth table for the material conditional modus tollens is valid because the only assignment which satisfies both premises ($p \rightarrow q, \neg q$) *is when both* p and q are false, in which case $\neg p$ is true. However, for the defective truth table there is no assignment to p and q which satisfies both the premises. This is because the conditional is now only true when p and q are both assigned true, in which case $\neg q$ is false. This does not represent a knock down argument, it simply points out that these may be insufficient grounds upon which to establish the desired conclusion that there is an asymmetry between the syntactic and semantic characterisations of a connective.

However, there may be further grounds on which to question the coherence of the distinction to which Johnson-Laird's argument appears to make appeal. The distinction in question is between syntax and semantics. Although this distinction has a clear basis in say natural language processing it may not be the case that the distinction is as obvious when it comes to defining a logic. The first question that needs to be raised is what is a logic? Johnson-Laird appears to define a logic as a set of syntactic rules for constructing derivations. Semantics is distinct in the sense that it describes the objects and relations which can be assigned to the syntactic formalism. Semantics is about the world. In this sense there is a clear distinction which is insisted upon in most textbooks. However, psychologically we are concerned with issues of process rather than the distinction between the language in which the world is described and the described world. The contrast between syntax and semantics is meant to capture the distinction between syntactic procedures for inference and semantic procedures for inference, not the distinction between a language and what it describes.

In more recent accounts, a logic is defined as an abstract *consequence* relation (Scott, 1971). That is, pairs of formulae (or sequences of formulae) of a language can be abstractly characterised as standing in a relation, which is designated a consequence relation, ie. $\phi \models \psi$. Two questions that can then be addressed. First, can a procedure be defined for deriving formula ψ from formula ϕ. Second, is their an efficient implementation of that procedure once defined. It is a demonstrable theorem of the theory of symbolic computing machines that any program is equivalent to some logical derivation (Howard, 1980). However, the efficiency of the algorithm will depend on what kind of procedures are adopted. So, let us return to the first question. The consequence relation can be characterised either syntactically or semantically. However, the *procedures* defined are equally formal, ie. they are defined over the symbols of the language. Whether t's or f's are being assigned to atomic formulae or syntactic inference rules applied, bears not one jot on the question of the formality of the operations. The question that can now be posed is whether Johnson-Laird's argument makes sense in this framework. The argument seems to amount to the claim that although there may be many semantic characterisations of a consequence relation, there can be only one syntactic characterisation. This claim is false. There can be many different syntactic characterisations of a consequence relation just as there can be many semantic characterisations.

The first thing to note is that whether characterised semantically or syntactically the logic is given by the consequence relation, not by its characterisations. Semantic proof procedures are as logical in this sense as syntactic procedures. It is perhaps a counter-intuitive product of this conception of a logic that soundness and completeness proofs are not establishing an equivalence between syntax and semantics. Rather any characterisation of a particular consequence relation can be sound and complete with respect to another. So two semantic characterisations could be sound and complete with respect to each other, further blurring any useful distinction between syntactic and semantic *processes*. Second, the statement of the truth conditions of a conditional for example, does not immediately constrain the set of inference rules which are included in a syntactic characterisation. For example, on the semantics for the material conditional, modus ponens must be included in the syntactic characterisation, only if the deduction theorem, ie. if $\phi \models \psi$, then $\models \phi \rightarrow \psi$, holds for the particular consequence relation. But syntactic proof theories may be defined which do not include modus ponens, while retaining this semantic characterisation of the conditional. This will mean that to syntactically derive an inference which accords with modus ponens may entail a rather more complex derivation than when modus ponens is included. This leads back to the second question concerning the efficiency of one system of proof over another. This question is complicated by the fact that a particular pair consisting of a

language and a consequence relation may be isomorphic to another pair. However, they may not be equivalent with respect to the complexity of the derivations employed. Hence, although a particular syntactic characterisation, eg. natural deduction, yields derivations of thus and so length, there may be a more efficient syntactic proof theory which yields semantically the same set of valid conclusions but with less consumption of computational resources. This point will be raised again in the conclusions.

It would appear that when concern centres on issues of process, the asymmetry Johnson-Laird seeks is unavailable *a priori*. Nonetheless it may be available *a posteriori*, ie. the empirically observed complexity of human inferential procedures may be incompatible with syntactic methods of proof. However, as will be discussed in the conclusions, once the asymmetry is seen to be spurious, then the possibility of mimicking the empirically observed complexity by *either* semantic *or* syntactic procedures has to be admitted.

Summary

This chapter has located the concept of inference embodied in the competence model of the preceding chapters in a general classification of modes of inference. Constraints are the generalities which license information gaining eductive inferences. Some of the properties of this inferential mode were shown to be reflected in subjects conditional reasoning behaviour. It was also shown how some conditional reasoning tasks, where apparently falsificatory behaviour was observed, can be provided with a rational basis when the small surveyable domains over which the generalisations stated in the rules are taken into account. This observation serves to place these tasks in a different category to other selection task data and they will not be discussed again in this thesis. It was then argued that partial interpretation provides a rational foundation for the observation of defective truth tables which also renders unsound the strategy of falsification. Some preliminary comments were also made concerning the foundations of Johnson-Laird's semantic account of human reasoning.

This chapter concludes Part I, on competence. In Part II, the implications of the competence model of eductive and inductive inference for psychological performance will be outlined. This will begin in the next chapter by considering the data obtained on Wason's selection task using only affirmative rules.

Part II:

Performance

Chapter 5: The Affirmative Selection Tasks

5.1 Introduction

In turning to Performance in the following chapters certain psychological assumptions will begin to enter in to the discussion. Competence has implications for performance, over and above justifying the rational basis of behaviour. Specifically, to explain performance it must be assumed that the cognitive system manifests knowledge of the competence model. This does not mean however, that the competence theory is directly implemented in the cognitive system. Rather the principles upon which the system works respect the competence model and in this sense can be said to embody knowledge of the competence theory. A performance theory which (i) respects the competence model, and (ii) can be seen as causally responsible for the observed behaviour must embody the sources of information which the competence model implies are necessary to perform various inferences. It must also provide explanations of the processes involved in deploying those informational resources. The competence model of Part I takes the principle source of information which determines inference to be pragmatic contextual knowledge. Later on it will be suggested that these sources of information are unlikely to be deployed and utilised in inference at the level of conscious decision making. The requirement to introduce some assumptions about the device indicates the need to distinguish the position to now be developed from the competence models of Part I, so it will now be given a name: Pragmatic Context Theory.

The results from Wason's Selection Task fall roughly into three conceptual groups:

(i) The abstract results.
(ii) The thematic "facilitation" results.
(iii) The Evans negations paradigm results.

Their order of presentation does not reflect their order of appearance in the literature, but it does reflect the recalcitrance of the results obtained to rational charaterisation. In this chapter the results obtained from Wason's selection task using only affirmative rules will be analysed within the framework of Part I. The basic abstract results will be dealt with first. This will include a discussion of the various *therapy* experiments (Wason, 1969, Wason & Johnson-Laird, 1970). The results from *thematic* versions of the task, which use "contentful" material will then be discussed. However, Part II will begin with a presentation of two examples from the last chapter which will be worked through in detail in the context of the four card problem. This will render explicit the assumptions which Pragmatic Context Theory relies upon in accounting for the data on Wason's task. These worked

examples will be appealed to later on in accounting for the actual data.

5.2 Examples

5.2.1 Non-taxonomic case: My hall light

To example the strategy of inquiry which would result from a non-taxonomic constraint the example of my hall light arrangement will be schematised. There are some rather obvious artefactualities about this example. Nonetheless, it functions to highlight the important features of the account and moreover makes the artefacts obvious.

The world

The situation of my hall lights can be described as a *disjunctive* Boolean function, ie. a function of two arguments *f*(A,B), corresponding to the two switches, each is a binary variable 1 = ON, and 0 = OFF. Let the domain of the function {<0,0>, <0,1>, <1,0>, <1,1>} be denoted l_1, for space time location 1. The range of the function {0,1} are the states of my hall light (L), let these be denoted l_2, for space time location 2, l_2 follows l_1: $l_1 < l_2$. By case the function is a disjunction, ie.:

(5.1) *f*(A,B) = 0 (ie. L = 0), if A = 0 and B = 0, else *f*(A,B) = 1.

This does not fully describe the situation in my hall since there are various conditions which must hold like the electricity is on, the bulb works etc. This can be added, making a function of three arguments, the third value is false if any one of the background conditions is false (ie. they are conjoined). This variable will be called C; the function is then:

(5.2) *f*(A,B,C) = 0, if A = 0 and B = 0, OR C = O, else *f*(A,B,C) = 1

This situation was described situation theoretically in the last chapter, I repeat that in full below.

Situation theoretic description

The constraints and eductions holding between the light switches being turned and the lights going are described as follows.

T_1: *<<Turn_light_switch*, l_1, x; *1>>* & *<<Switch_A*, l_1, x; *1>>*

T′: *<<Light_on*, l_2, y; *1>>* & *<<My_hall_light*, l_2, y; *1>>* & *<<Follows*, l_2, l_1; *1>>*

$C_1(\Psi S_1)$: <<[n]=>, T_1, T′, T″; *1*>>/ ΨT_1 --> $\Psi T' \mid \Psi T''$

The constraint holding between light switch B being turned and the light going on simply replaces T_1 with T_2:

T_2: *<<Turn_light_switch*, l_1, x; *1>>* & *<<Switch_B*, l_1, x; *1>>*

$C_2(\Psi S_2)$: <<[n]=>, T_2, T′, T″; *1*>>/ ΨT_2 --> $\Psi T' \mid \Psi T''$

The types in these constraints are clearly very local and non-general, the generality is given by the space time location parameters.

Each of C_1 and C_2 potentially has a corresponding *converse* constraint and eduction:

$CC_1(\Psi CS_1)$: <<[n]=>, T′, T_1, T″; *1*>>/ $\Psi T'$ --> $\Psi T_1 \mid \Psi T''$

$CC_2(\Psi CS_2)$: <<[n]=>, T′, T_2, T″; *1*>>/ $\Psi T'$ --> $\Psi T_2 \mid \Psi T''$

Assumptions

Now some assumptions:

(1) A crucial assumption is that the inquirer is ignorant of what precisely the set up is. He is, after all inquiring into the nature of his world, he has no God's eye perspective.

(2) Moreover, his strategy will be an open system strategy, the domains of the location parameters are not tied to only those articulated in the examples.

(3) The predictive cycle strategy indicates that subjects should attempt to make eductions using the constraint. Each situation that will be considered is *independent* of the others. This is the normal situation an inquirer finds himself in. In any given situation he will attempt to gain information the best way he can. However, in each new situation he can not

be certain that the same background conditions hold or whether other constraints, perhaps to which he is not attuned are operative. Since the four situations to be considered correspond to the the four cards in Wason's task, it is appropriate to observe that the card selections have been demonstrated to be statistically independent (Evans, 1977).

(4) For a non-taxonomic constraint, the indicating and indicated types describe *discrete* events. Therefore the cards can not be conceived of as *instances* of unified objects to which properties are being assigned. This is carried over to the cards in the selection task where each side is taken to represent a different space time location (cf. below)

(5) The constraint being investigated is C_1, I denotes the inquirer.

The four situations

The four possible situations, corresponding to the four card, in which an inquirer may be in various states of ignorance about the system of lights are illustrated in Fig. 5.1. For each which eductions are sound and for what reasons will be outlined.

P-card In this situation s_1 at l_1, A = 1. This means there is an anchor f for I_1: f: I_1 |-> l_1. Since A = 1, the proposition that s_1 |= $T_1[f]$ is true. This leads to the eduction (ΨS_1) that at l_2, s_1 |= T'[g], where g extends f such that: g: I_2 |-> l_2. That is, at the time of observing that light switch A is turned I (our inquirer) uses ΨS_1 to educt to the expectation that at a later time l_2, the light will come on. He does this in ignorance of whether the background conditions hold since he is ignorant if any actually do hold, a predictive failure would initiate inquiry in to what they were.

¬P-card In this situation s_2 at l_3, A = 0. This means that although there is an anchor f for I_1: f: I_1 |-> l_3. Since A = 0, the proposition that s_2 |= $T_1[f]$ is false. Since this is the case perhaps I could educt to the falsity of s_2 |= T'[g], where g extends f such that: g: I_2 |-> l_4. The grounds for not making the specified eduction is the general assumption that other constraints, eg. C_2 could be operative. If this is the case, then if B = 1, then the eduction (ΨS_2) to L = 1 could be made (by parity of reasoning with **P-card**). Since this is the case eductions which assume T_1 to be necessary for T', are unsound. Subjects in the task are told to only turn the cards they must, ie. only those which are licensed by the constraint, in this situation ΨS_1 would not be expected to yield information.

For both the antecedent card situations the predictive constraints C_1 and C_2 have played a

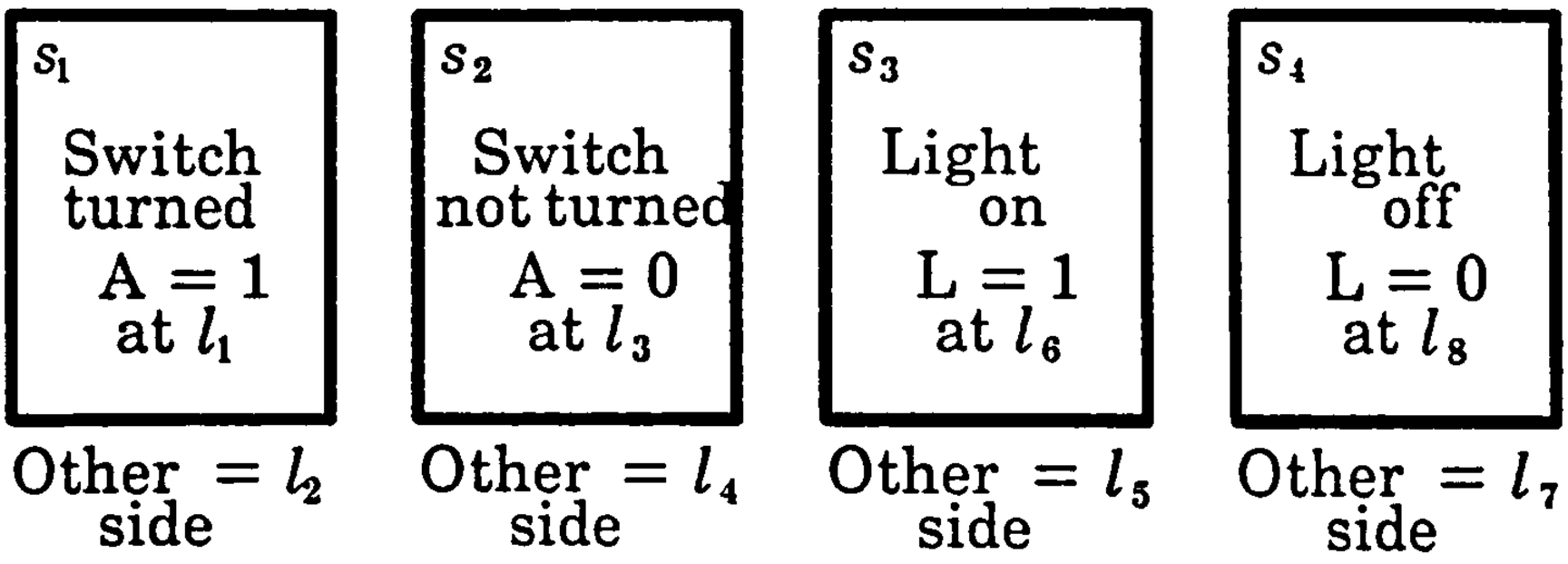

Figure 5.1: My Hall light: The Cards

role. When predicting future events which track the natural causal/action/temporal order, these are the only constraints relevant and of themselves they license no eductions in the converse direction. For the consequent cards, subjects are performing a *different task*, they are having to ask whether in a situation where they know something about what happened at a later time, whether they can *explain* this event. (This distinction is borne out in the psychological literature where data on the consequent cards is often equivocal. Manktelow & Evans (1979), for example, hypothesize that this is due to only subjects' antecedent card selections being logically determined.) However, this involves utilising the *different* converse constraints. Generally, the 33% of subjects (cf. chapter 4 & Table 5.1, below) who don't turn any consequent cards are attending to the initial task requirements more diligently than subjects who take their role to be to see whether explanations for these events can be derived. However, many subjects may well assume that the mere presence of the consequent cards indicates that they should seek explanations. Again in considering these cards, it will be shown how different contextual assumptions may drive differing selections.

Q-card In this situation s_3 at l_6, L = 1. This means that there is an anchor f for l_2: f: l_2 |-> l_6. Since L = 1, the proposition that s_3 |= T'[f] is true. Since this is the case perhaps I

could use ΨCS_1 to educt to $s_3 \models T_1[g]$, where g extends f such that: g: $l_1 \mapsto l_5$. I is now attempting to explain the light being on. If he assumed that there are no other constraints operative which could also possess corresponding converse constraints, eg. CC_2, then he could educt to A = 1. However, if he assumes other constraints are operative, namely CC_2, then that L = 1, only allows the eduction (ΨCS_1 & ΨCS_2) to A = 1 or B = 1 (note that since L = 1, it must be the case that C = 1). So, even if I were to assume that he should attempt to explain the consequent card events, he would still have grounds not to turn the *q*-card. However, if he also assumed B = O then again he could educt using ΨCS_1 to A = 1. In general, the results (cf. below), indicate that subjects do tend to assume they can make this eduction.

¬Q-card In this situation s_4 at l_8, L = 0. This means that although there is an anchor f for l_2: f: $l_2 \mapsto l_8$ since L = O, the proposition that $s_4 \models T'[f]$ is false. Since this is the case perhaps I could use ΨCS_1 to educt to the falsity of $s_4 \models T_1[g]$, where g extends f such that: g: $l_2 \mapsto l_7$. The grounds for not making this eduction are that in this situation I can not assume that C = 1. When L = 0, either A = 0 and B = 0, *or* C = 0. Since this is the case this eduction could not be expected to allow information gain. I is attempting to explain the fact that L = 0. However, inquirers do not normally attempt to explain the non-occurrence of all the events which are not occurring at any given time. In I's purview most things are not happening. It only makes sense to attempt to explain the non-occurrence of an event if *it was predicted to occur*, as in the predictive cycle. However, if L = 0 represents a predictive failure, then switch A will have been turned!, ie. A = 1, and concern will centre on which background condition failed.

This example articulates the grounds, relative to general context dependence assumptions, upon which various eductions are performed and why others are not when a constraint is non-taxonomic. When looking at subjects actual task behaviour it is important to realise that the possibilities made explicit above are not intended to function as hypotheses concerning possibilities subjects actually consider. They are possibilities not explicitly ruled out by standard versions of the task. In describing the tasks below examples of possibilities left open analogous to those above will be mentioned. However, their function is purely illustrative of possibilities always left open in peoples inquiry into the real world. The strategy attributed too subjects in the task and which license the selections they make, is a product of their normal strategies of inquiry. Unless, as in the various COSTs, additional circumstantial features elicit different interpretations and hence strategies they will default to the predictive cycle strategy.

There is a very obvious degree of artefact in the above example. The problem with all toy models is that, by their very nature, they tend to make the simplifying assumptions which the account being developed here explicitly wants to avoid. The choice of example, was conditioned by the need to give a precise and exhaustive description. The problem with this is that when people are in possession of such a description they are able to reason about the situation in ways they cannot when dealing with the real world. The limited number of possible states and operative constraints, makes it very tempting to argue that other inference patterns are licensed. Although, the reader has been exhorted to adopt I's point of view, ie. *within* each situation and ignorant of the state of the world, once you *can* assume a God's eye perspective it very hard not to do so.

5.2.2 Taxonomic case: Johnny and the pipes

The same basic schema will be used here as for the last example. Presenting this taxonomic example serves to illustrate the various strategies introduced in chapter 4 in discussing the COST and to provide the motivation for an ambiguous result found in the thematic versions of the selection task, due to Wason & Shapiro (1971).

The world

Providing a description of the set up of the world is less complex here. There is a box of pipes of various dimensions, some threaded some not, some are also bends. Let P be the set $\{P_i | P_i$ is in the pipe box$\}$. Certain subsets of P will prove of interest, ie. the set *1"*: $\{P_i | P_i$ is 1"$\}$ and the set *Th*: $\{P_i | P_i$ is threaded$\}$. In working through the different strategies available another dimension of the sets will prove relevant, ie. their size or *cardinality*, specified, eg. $|P|$. Apart from the claim that, $|P|$ is small (in the case of the four card example it will be 4), it will only be the *relative* cardinalities of the various sets which will be of concern. It could be the case that P is a sample of a larger population P, ie. $P \subset \mathbf{P}$.

Situation theoretic description

The only constraints operative are as follows:

T_1: $\ll$*1"_pipe*, **x**; *1*$\gg$

T': $\ll$*Threaded*, **x**; *1*$\gg$

C_1 <<[n_t]=>, T_1, T′; *1*>>

and

T_2: <<*Bend*, x; *1*>>

C_2: <<[n_t]=>, T_2, ¬T′; *1*>>

Note relative to the treatment of duals ("¬") in chapter 2, the negation in C_2 carries the information that it has a smooth bore and this precludes being threaded, ie. in this context threaded and unthreaded are antonyms.

Assumptions

(1) One point of contrast between taxonomic and non-taxonomic constraints is that the former provides a unified object to which various properties are being assigned. This allows a coherent interpretation in terms of instances (cf. the discussion of Wason & Green, 1984). This renders coherent the concept of an objectual restriction on the domain of objects which makes less sense in the non-taxonomic case. This also works for unifiers which are relations, for example, "if I travel to Manchester I take the train" is objectually restricted to my travelings to Manchester in my life time. Or it may just be known that I only ever traveled to Manchester 4 times.

(2) Given a closed domain of instances, allows the prior assessment of the force of each instance. The relevant situation is given by the box of pipes. Johnny, can not conceive of each instance as an isolated opportunity to use the constraint to make an eduction. Rather given the situation, the pipe box, he must determine relative to knowledge of the prior distributions (or a lack of such knowledge) which strategy to adopt in determining the truth or falsity of the constraint.

(3) A taxonomic constraint encodes a limited type-token relationship between indicated and connecting type. The indicated type is a restriction on the appropriate anchors for one or more parameters in the indicating type. Relative to peoples natural taxonomies this relation may be obvious, eg. All ravens are black. However, situations do arise, like Johnny's, where there is uncertainty concerning the relative sizes of the domains. This is precisely why the strategies appropriate alter, when different background knowledge about those domains changes.

(4) A similar cautionary note applies to this example as to the taxonomic case. The

strategies that Johnny may adopt are ones natural to his situation given the various background knowledge he may or may not possess. On performing a selection task the most obvious, real world strategy is to turn all the cards. However, the exhortation to turn only the cards you *must*, indicates that only those licensed by some other more reasoned and parsimonious strategy of inquiry, should be turned. The example is worked through on the assumption that subjects are importing strategies which they would use in real situations.

When Johnny is given his instructions, he must use his prior knowledge or lack of it to determine the appropriate strategy. This simply involves determining what kinds of instances he wants to look at. The situation analogous to the 4-card problem is illustrated in Fig. 5.2. The analysis will not be presented card by card but strategy by strategy. For each, the cards he should turn will be described. For each strategy the background knowledge which licenses it will be initially described.

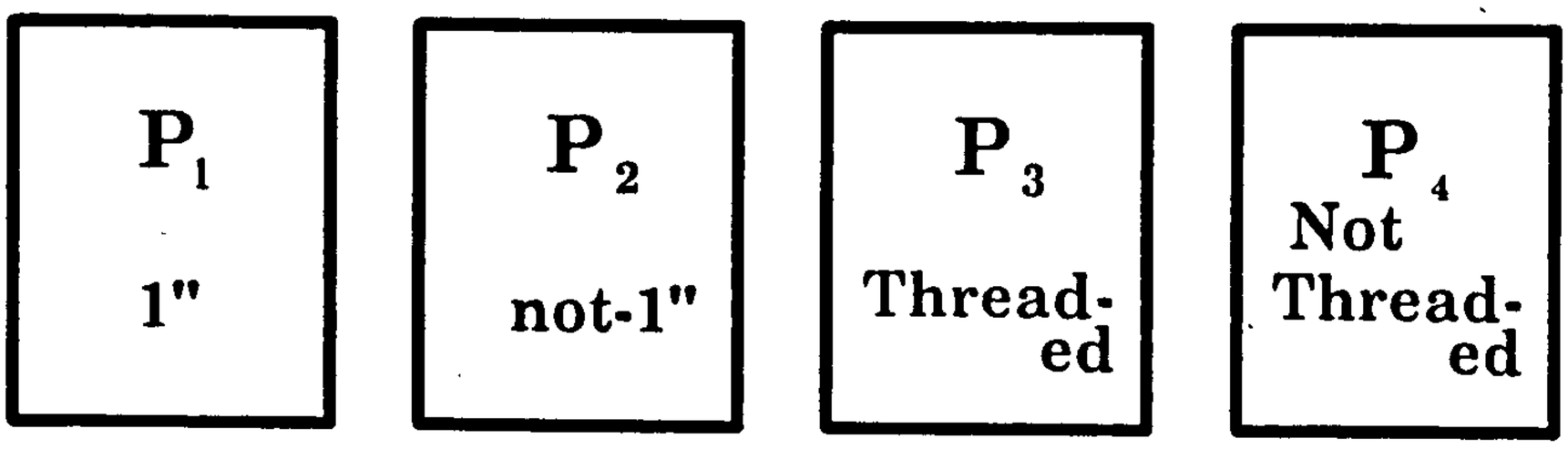

Figure 5.2: Johnny and the Pipes: The Cards

Closed Domain

1. Positive Confirmation: Johnny knows that $|Th| = |Th'|$, where "′" is the contrast class operator, ie. Th' is the set of smooth pipes. However, he also knows that there are many other sizes of pipes in the box and hence it is most likely that $|1"| < |1"'|$. Given this information, it follows that it is most likely that $|1"| < |Th'|$ by some orders of magnitude, thus looking exhaustively for 1" pipes to check they are threaded has the most labour saving potential. In this case he would select only P_1. However, the limited situation provided by the cards means that there is only one instance such that $P_i \in Th$, ie. P_3. Generally, since $|Th| = |Th'|$, looking at threaded pipes as well as 1" pipes would prove labour intensive. So, he would also turn P_3.

2. Negative Confirmation: Johnny knows that $|Th'| < |Th|$, ie. most pipes are threaded. However, he is ignorant of $|1"|$ and hence $|1"'|$. He, therefore looks for unthreaded pipes to check they are not one inch, so he would turn P_4. However, lest the rule be vacuous, he also checks P_1. What would happen in this case if Johnny found that $P_4 \in Th'$, ie. it was unthreaded? Well if he also discovered that it was a bend, and he was told that C_2 is operative, then even this would not falsify. Only if it can be assumed that P is not subdivided in other ways which are relevant to his determination of the truth of C_1, will his discovery of a "falsifying" instance, retain its falsificatory status. Even in an objectually closed domain this can not be guaranteed and in an open domain it is a positive expectation that this is unlikely to be the case.

Open Domains

In an open domain, where Johnny knows that the pipes in front of him are simply a sample of a much larger population, the characteristics of which he is unsure of, then the normal predictive cycle strategy will apply. If he discovers any unthreaded 1" pipes then this will initiate inquiry into whether, for example, C_2 holds.

The artefact in this example, concerns the fact that in normal inquiry another factor is important which may affect the strategies adopted. This is the *ease of property determination*. Along with the respective domain sizes, the relative ease of determining whether an object possesses a certain property is important. For example, suppose that the various dimensions of the pipes are quite close, and therefore a measuring instrument has to be employed. However, the "threadedness" of the pipes is easily determined, eg. they are external threads. In this situation, even given the relative domains which license positive

confirmation, it would be an unreasonable strategy to adopt. People need to adapt their information gaining strategies such that they are appropriate to *effective* and *efficient* action. This issue will be raised again in discussing the thematic results.

The basic abstract results will now be provided with a rational grounding in the same terms as the two examples.

5.3 The abstract results

The standard paradigm for these experiments was described in chapter 4. For perspicuity, table 5.1 below shows the standard results. Over the years there have been many permutations of this basic abstract result with the logically correct p, $\neg q$ selection sometimes rising as high as 19% (present experiments). However, for the standard paradigm without any further manipulations table 4.1 represents the data to be accounted for. The only manipulations which had a facilitating effect, other than certain therapy conditions and various COSTs (which have been discussed above), will not be dealt with until the section on the thematic facilitation results.

Providing an account of the data will involve attributing various sources of information to the subjects. These are not to be thought of as consciously available to subjects in the course of their task performance. The contextual information which enters into their final interpretation is processed at an automatic and preconscious level. Certain results of the process are of course available to consciousness, but the basic mechanisms involved which muster the information sources determining a particular interpretation are not. This is a wholly familiar view of the processes involved in parsing which are not assumed to be available to conscious introspection, although the products of the process are (Marslen-

Table 5.1
Standard Abstract Selection Task Results
(Johnson-Laird & Wason, 1970)

p and q cards	46%
p card only	33%
p, q and $\neg q$ cards	7%
p and $\neg q$ cards	4%
others	10%

Wilson & Tyler, 1980; Crain & Steedman, 1985; Wanner & Maratsos, 1978). It is also assumed that in standard selection tasks subjects interpret the rule as applying to an open system (cf. above). In which case other contextual assumptions may be bought to bear on interpreting the rules which influences the inferences subjects take the rule to license. Those inferences (eductions) will in turn influence the preconscious processes of belief fixation embodied in the predictive cycle. This follows Wason & Evans (1975) and Evans (1980a, 1980b) interpretation of the protocol data. Interpretation and confirmation strategies are processes which in the first instance are preconscious and automatic. The products of these processes can be made available to consciousness, but their primary function is to extract information appropriate to guiding action. This establishes points of direct comparison between the theoretical account being developed here and Wason & Evans' dual process theory. In the next two chapters the positions will be contrasted over the issue of the nature of preconscious processing which Evans (1972, 1973, 1982, 1983) views as relatively insensitive to interpretational factors.

For the standard abstract version, an exhaustive analysis will be provided on the same lines as the examples. The purpose is to establish the consistency of the competence account with the actual data. This will only be done once. For each card an account will be provided of why a subject would be expected to make the selections which are reflected in the *modal* responses observed in the data. Where a rational basis for other responses seems forthcoming, and hence need not be put down to pure error, they will also be mentioned. The situation theoretic interpretation of the rule will first be provided along with some elucidatory comments, then the grounds for selecting each card will be outlined. Often a *background* rule is provided in these task, in the standard task it was as follows:

(5.20) There is an number on one side of each card and a letter on the other side.

The cards employed on the task showed the following letters and numbers:

A(p) K(¬p) 2(q) 7(¬q)

And the foreground rule was as in (5.21)

(5.21) If there is a vowel on one side of the card, then there is an even number on the other side of the card.

This rule is ambiguous between a stating a contingent taxonomic constraint (cf. (4.16)) (the corresponding eductions will be assumed for this discussion):

T_Σ: *<<Card, one-side:* **x**, *other-side:* **y**; *1>>* & *<<Vowel,* **x**; *1>>*

$T'_{\Sigma'}$: <<*Even_number*, y; *1*>>

(5.21′) <<[c_t]=>, T_Σ, $T'_{\Sigma'}$; *1*>>

or a non-taxonomic constraint:

T_Σ: <<*One_side:* x; *1*>> & <<*Vowel:* x; *1*>>

$T'_{\Sigma'}$: <<*Other_side*, y; *1*>> & <<*Even_number:* y; *1*>>

(5.21′′) <<[c]=>, T_Σ, $T'_{\Sigma'}$, $T''_{\Sigma''}$; *1*>>

Unless prior beliefs can be appealed to in order establish the *unity* of instances, the disjoint non-taxonomic interpretation predominates (cf. Wason & Green, 1984). Subjects assume that that are confirming in an open domain (cf. Beattie & Baron, 1988). Therefore, they treat the cards as a sample of a much larger population. Hence, (i) subjects will adopt the open system predictive cycle strategy, and (ii) they will not assume that the domain is fixed in advance, other predicates may apply to alter the context of the generalisation, and other constraints may be operative determining the distribution of letters and numbers on the cards. These assumptions are not explicit, they simply reflect the subjects normal epistemic relationship to the open system which is their world.

The model of the predictive cycle predicts that subjects should use the rule/constraint in order to see if it permits information gain. ie. use it to predict what is on the other side of the cards. Constraints can license eductions in either direction, the predictive or the explanatory. However, the circumstances which allow unrestricted eductions in the explanatory direction are different. If an assumption concerning the existence of other predictive constraints is unwarranted, then subjects may consider it appropriate to check whether the constraint licenses an eduction in the explanatory direction. The foreground rule does not explicitly mark any reversal of causal/action order, nor does any prior pragmatic world knowledge suggest such a reversal. Therefore, subjects must assume the antecedent-consequent order tracks the predictive causal/action order.

Each card can be thought of as a situation *s* in which the subject is attempting to discover whether the constraint holds. It should be understood that in the non-taxonomic case each situation is treated independently. Each situation presents an opportunity for a subject to make an eduction using the rule, ie. to test its predictive utility.

A-card The constraint states that even numbers can be predicted from vowels. Since A is a token of the type vowel, subjects select this card to see if a predictive success results. This

is required even in ignorance of whether or not any background information needs to be fixed to allow successful prediction. Identifying those conditions is only initiated by predictive failure.

K-card Situations in which the indicating type of a constraint does not hold bear neither one way or the other on the truth or falsity of that constraint. This is a function of partial interpretation. Hence the K-card is not turned.

2-card (5.21) has a corresponding explanatory constraint which suggests the possibility of educting from an even-number to a vowel (E => V). The actuality of this constraint E => V, does not bear directly on the actuality of (5.21), although it does bear on its *explanatory utility*. Behaviour on this card is also mediated by an ambiguity in interpretation dependent on which contextual assumptions are bought to bear. If other perhaps more specific constraints are operative, eg. x => 2, then subjects would be unlikely to turn the 2 card, since whatever they find will not bear on the original constraint. So, there is a division between subjects:

(i) The majority, assume that explanatory utility bears on the information gaining potential of the original constraint. This functions to detract attention from the possibility that other constraints may be in operation. However, if they turn the 2-card, whatever they find bears neither way on the original constraint.

(ii) The minority, assume that other constraints may be in operation. And that therefore whatever is on the other side bears neither way on the original constraint. These subjects do not turn the 2-card.

7-card All constraints are context sensitive. Hence, as has already been extensively argued, explanatory eductions from the non-occurrence of the indicating type (7) to the non-occurrence of the indicated type (odd-number) are unsound. Moreover, inferences from the occurrence of the indicated type without the indicating type (predictive failure) to the falsity of a constraint (general rule) are similarly unsound. Therefore, subjects need not turn the K-card. It may well be that on turning this card a subject discovered a "y", and concluded the rule was false only to discover that the experimenter had not intended "y" to be included in the set of vowels.

Some early manipulations of the task could have been predicted to fail in their goal of eliciting falsificatory behaviour. Wason (1969) focused on the possible confusion bought

about by the fact that conditionals in natural language are often used to encode information concerning either a causal or temporal relation holding between antecedent and consequent. "Hence, the simultaneous presentation of values of them could have induced considerable perplexity" (Wason and Johnson-Laird, 1972:176). In an attempt to avoid this, the rule was changed to "Every card which has a red triangle on one side has a blue circle on the other side". No facilitation was found. One plausible reason for this was that subjects were interpreting "one side" in the antecedent as the upwardly turned face and in the consequent as the downwardly turned face. Consequently, Wason and Johnson-Laird (1970) altered the task materials such that all the information was on the upwardly turned face. Each card had a shape in its centre and a number of borders. The partial nature of the information available to the subjects was retained by the use of masks. The rule retained the universally quantified form used in the above experiment: "Every card which has a circle on it has two borders round it". No facilitation was found for this variant of the task, and the results simply reflected those obtained in the standard form.

The procedure used by Wason and Johnson-Laird (1970) only eliminated the upward/downward directionality, but not the known/unknown directionality. Eliminating the latter would, of course, destroy the task. Moreover, if the upward/downward directionality was the critical factor then one would expect the *p* card only as the modal response in the standard form, which was not observed. Since the crucial directionality involved in the task is the known/unknown directionality, which requires the subjects to perform an eduction, this manipulation would have been predicted not to facilitate falsificatory responding.

5.3.1 The therapy experiments

The first manipulations to produce a notable facilitation were introduced in the "therapy" experiments of Wason (1969) and Wason and Johnson-Laird (1970). The basic idea of these experiments, as their label suggests, was to provide therapeutic procedures to enable subjects to see where they were going wrong without giving the whole game away. The method was to engage subjects in a dialogue concerning their task performance exposing them by degrees to inconsistencies "between their initial selections of cards and their subsequent independent evaluations of specific cards as falsifying or verifying the rule" (Wason and Johnson-Laird, 1972:179). These experiments were not wholly successful, even when subjects were subjected to the most concrete inconsistency, ie. turning the $\neg q$ card to reveal a *p* on the other side, only 42% got it right although 78% now selected the $\neg q$ card as opposed to only 6% in the initial condition. However, certain graded effects were observed

which Wason and Johnson-Laird (1972) subsequently incorporated into their "insight" model.

Two such models were proposed by Johnson-Laird and Wason (1970). The first reflects the order of responses in the standard results, and attempts to explain them in terms of two forms of "insight". Lacking either form, subjects are postulated to attempt verifying the rule which leads them to select only the p and the q cards. Possessing insight (a) only, leads subjects to only check those cards which could verify to see if they could falsify, ie. they select only the p card. Possessing insight (b) only, leads subjects to select those cards which can verify *and* of those cards which do not verify, those which could falsify, ie. they select the p, the q and the $\neg q$ cards. Possessing both, will lead to the correct selection of the p and $\neg q$ cards, as the two insights are assumed to be additive. Wason and Johnson-Laird (1972:183) point out that this model "is almost certainly grossly wrong". Failing to select $\neg q$ (lack of (b)) occurs far more frequently than selecting q (lack of (a)). They observe that although the two errors have the same logical status they may have different psychological sources. And they surmise that the selection of just p "may not be the result of the (non-trivial) insight that verifying cards should be rejected if they could not falsify, but *merely* (my italics) signifies that the subject does not assume that the converse of the rule holds" (Wason and Johnson-Laird, 1972:185). The therapy experiments also enhanced the implausibility of the original model. Hypothetical contradiction, where the subject simply says that p on the other side of the $\neg q$ card would falsify, induced a switch in selections from p only to p, q and $\neg q$. This corresponds to a simultaneous loss of insight (a) and gain of insight (b), which as they observe is "psychologically implausible and bizarre". Gaining insight (a) alone, ie. switching from p and q, to just p was also very rare in the therapy experiments.

The revised model accounted for these observations and results as follows. Subjects were assumed to initially focus only on the cards which are mentioned in the rule. Dependent on whether they believe the converse to hold or not, they select only those cards which could verify, ie. p and q or p only respectively. Two interdependent levels of insight then determine subsequent responses. Possessing *partial insight*, leads subjects to test all the cards selecting those which could verify *and* those which could falsify, ie. selecting p, q and $\neg q$. Possessing *complete* insight, consists in the realisation that only cards which could falsify the rule need to be selected.

As observed by Bree (1973) and reported in Evans (1982), there is an inconsistency in the revised model. The claim is that subjects initial no insight choices are determined in two

stages, (i) look only at cards mentioned in the rule, (ii) verify under the constraint that either the converse holds (turn *p* and *q*) or only the implication holds (turn *p* only). Admitting the converse is equivalent to treating the rule as material equivalence, which, when in a state of partial insight, should lead subjects to turn all four cards, a response barely ever observed. This observation lead to several modifications of the insight model (Bree and Coppens, 1976; Smalley, 1974). On Smalley's model, which incorporates the defective truth table (Wason, 1966), subjects either initially adopt a defective implication or equivalence interpretation, and subsequent responses are determined by whether they perceive the reversibility of the cards or not and what level of insight they attain.

Evans (1982) criticises these models in part because of the protocol data upon which they are partly based and in part because they fail to deal with results he obtained when incorporating negative components into the rules. The latter will not be dealt with until the chapter on the Evans' negations paradigm. My own criticism centres on the non-explanatory invocation of the concept of "insight" in these models. "Insight" as it originally appeared in the Gestalt literature (eg. Wertheimer, 1945; Duncker, 1945; Koffka, 1935; Luchins, 1942) referred to a particular phenomenon: "the famous "Aha!" experience of genuine (understanding)" (Wertheimer, 1985) not an explanatory construct. The "insight?" question diamonds in the flow charts used by Wason and Johnson-Laird (1972) to illustrate these information processing models should be replaced by clouds indicating processes as yet undefined or unknown (Colin, 1980:73). The sub-processes which would have to replace the clouds would need to involve how people represent the task and how those representations change in order to generate the representations and operations appropriate to solving the task correctly (Wertheimer, 1985). It is this productive generation of representations and operations appropriate to solving, or partially solving, a problem which the Gestalt psychologists originally felt captured the concept of "insight". The models do not deal with *representational* issues but outline a computational *process* for producing the appropriate responses, assuming people interpret the rules and cards in various different ways. A full explanation would have to outline what is represented and how, and also how those representations and/or their content altered so as to produce the various stages of insight. In providing a rational basis for this data attention will concentrate on articulating the changes of content which could occur in response to the therapies and which would function to alter subjects' performance.

The modal transitions described below come from Wason (1969). Two responses predominated in the initial choices: *p* card only (50%) and *p* and *q* card (41%). These responses constitute the baseline from which it was hoped the therapies would produce some

alterations in performance. Pragmatic context theory has the virtue of connecting subjects' initial responses to their subsequent transitions in inferential behaviour. The main transitional sequences were:

(I) $p \Rightarrow p,\ q\ \&\ \neg q \Rightarrow p,\ \neg q.$

(II) $p,\ q \Rightarrow p,\ q\ \&\ \neg q \Rightarrow p,\ \neg q.$

The secondary sequence was very rarely completed. Out of the 15 subjects who finally adopted the "correct", $p,\ \neg q$, response, only 3, chose $p,\ q$ initially, whereas, 10 chose p only initially. $p,\ q \Rightarrow p$ transitions were very rare, and the modal transition from an initial choice of p only was, $p \Rightarrow\ p,\ q\ \&\ \neg q$, indicating the scarcity of $p \Rightarrow\ p,\ q$ transitions.

The materials used in Wason (1969) were coloured shapes. These materials produced no significant alterations in initial performance over the standard version. The rule interpretations are the same as the non-taxonomic case above (5.21′), with appropriate substitutions. All the assumptions made in accounting for the modal responses in the standard task are also operative. Each sequence is dealt with separately.

Sequence I These subjects begin with the assumption that the system is completely open and hence all contextual factors could be operative (cf. card by card analysis above). In response to the therapies, they first move to an interpretation involving no contextual assumptions being in force, ie. they can educt freely in the explanatory direction. This leads to them to make eductions from q to p, and from $\neg q$ to $\neg p$ (cf. above). In response to further therapies, these subjects, then make the transition to assuming that there may be other constraints in operation, which prevents eductions from q. This entails admitting some contextual assumptions concerning the possibility of other constraints involving the q card, while dismissing another: the general context dependence of the rule. Making assumptions which allow one contextual variable to be operative while dismissing another, may be responsible for the fact that only 42% of subjects finally made falsificatory responses. However, sequence I subjects have already conceded that other constraints may be operative in their initial selections (p card only), so it may be reasonable for them to assume that they are supposed to now reject that assumption in response to the therapies. This contrasts with subjects in sequence II.

Sequence II These subjects begin by assuming that there are no other constraints operative which could invalidate an eduction from q to p. In response to the therapies, they then move to assuming that the rule is in general not context dependent and so they can educt from $\neg q$ to $\neg p$. However, these subjects began with the assumption that no other

constraints were operative involving *q*, they would, therefore, be less inclined than sequence I subjects to abandon this assumption and therefore adopt the set of contextual assumptions which yield a falsificatory response profile. This interpretation is borne out in the data where out of the 42% of all subjects who finally made falsificatory responses, only 20% where in sequence II while 67% were in sequence I. Initial selections were *p* card only, 50% and *p* and *q* cards, 41%. This indicates that the conditional probability of making a falsificatory response given an initial *p* selection was 0.56, the conditional probability of such a final response given an initial *p* and *q* card selection was 0.2, ie. subjects were 2.8 times as likely to make a final falsificatory response given an initial *p* card only selection than a *p* and *q* card selection.

Subjects' response to the therapies can be characterised as asking the question, "Have I got the circumstances right?". Their responses are determined by adjusting various contextual factors which, dependent on the initial assumptions made, will determine whether subsequent responses which accord with falsification are forthcoming. Note that the locus of the therapeutic procedures was wrong. Asking subjects to concede that *p*, $\neg q$ instances were false instances of the rule, could not be taken to promote a falsificatory strategy. *p*, $\neg q$ is a false instance of the rule but it does not necessarily falsify the rule. Since this is the case, the therapies would not be expected, of themselves, to elicit many falsificatory responses, as was observed. However, subjects have taken the therapies to have a purpose, and have made various interpretative adjustments. Pragmatic context theory makes clear predictions concerning the kind of manipulation which would be expected to elicit falsificatory responses, ie. explicitly manipulate subjects contextual assumptions.

5.4 The thematic facilitation results

The first experiment to use thematic material was conducted by Wason & Shapiro (1971). They used the rule which has been employed throughout the preceding chapters as an example of a relational taxonomic constraint, ie.

(5.10) Every time I go to Manchester, I travel by train.

In this experiment 63% of subjects made responses consistent with falsification. Early interpretations of this result centred on the ability of realistic content to access the appropriate logical rules, ie. to facilitate insight. Another early experiment used a postal rule and instead of cards, envelopes (Johnson-Laird, Legrenzi & Legrenzi, 1972). The rule was:

(5.10) If a letter is sealed, then it has a fifty lire stamp on it.

Rather than cards the task materials consisted of four envelopes, sealed, unsealed, one with a fifty lire stamp and one with a forty lire stamp. 81% of subjects made responses which accorded with falsification.

Several further studies leant additional empirical weight to the argument that realistic materials facilitated reasoning on the task (Bracewell & Hidi, 1974; Gilhooly & Falconer, 1974). These latter studies also attempted to separate out the determinants of this facilitation into either thematic content or the employment of a realistic relation, but with equivocal results. Manktelow & Evans (1979) presented the first data to seriously raise doubts about the validity of the thematic materials effect. In a series of experiments no facilitation was observed. Experiments 1 to 4 employed rules such as:

(5.12) If I eat haddock, then I drink gin.

They also systematically permuted negations in the rules, a procedure which will not be discussed until the next chapter. However, their failure to replicate was not a artefact of pooling the data across the rule forms, no facilitation was observed even on the purely affirmative rule. Importantly, in experiment 5, they also failed to replicate the original Wason & Shapiro (1971) study.

Although they do not refer to it as such, Manktelow and Evans (1979) were the first to broach the possibility that a significant proportion of the observed facilitation may be due to *memory cueing*. The rules used in many early experiments were in the specific experience of the subjects. For example, a similar postal rule was in force in Britain at the time the experiments were conducted. This could allow subjects to simply remember the appropriate instances without having to engage in any reasoning at all (Evans, 1982). This possibility was further tested by Griggs & Cox (1982), who using the postal rule found no facilitation for American subjects not familiar with the rule. However, facilitation was observed for an under age drinking rule:

(5.13) If a person is drinking beer then that person must be over 19 years of age,

a law that would have been familiar to the Florida state University students who acted as subjects, since the law was in force in that state at the time. A related observation made by Pollard (1982) and Pollard & Gubbins (1982) concerns the possible effects of context. In the under age drinking experiment, and Johnson-Laird et al (1972), subjects were provided with a context or scenario. This involves asking the subjects to imagine they are either postal workers checking mail or policemen checking for under age drinkers. It could be that

the main determinant of subjects' responding is a function of context rather than thematic content per se. Recently, Pollard and Evans (1987) have attempted to separate out the effects of content and context, using similar materials to Griggs & Cox (1982) they discovered that appropriate context *and* content is necessary for a facilitation of the falsificatory response.

The memory cueing hypothesis is also consistent with the observation of truth status effects (Pollard, 1979; Evans & Pollard, 1981). If a cued memory indicated a negative association between antecedent and consequent, then in accordance with belief bias subjects should look for falsifying instances. However, two results did exist which tended to raise doubts about the memory cueing hypothesis. First, the postal rule was used on British subjects who could not be supposed to have had experience with Italian postal regulations. Since a similar rule was in force in Britain, perhaps some reasoning was involved if only analogical rather than logical. Second, a rule used by D'Andrade (reported in Rumelhart, 1980), could not have been in the direct experience of subjects although it produced a marked facilitation. A scenario was presented telling subjects to imagine they were store managers checking receipts, the rule was:

(5.14) If any purchase exceeded $30, the receipt must have the signature of the department manager on the back.

Since the imaginary store was Sears, this goes by the name of the Sears version. Almost 70% of subjects made responses which accorded with falsification. However, the Sears rule and the under age drinking rule share a potentially important feature. Both relate to various obligations (or conventional constraints), ie. rules about which there is some independent necessitation, signified by the modal *must*.

Cheng & Holyoak (1985, 1986) develop this idea in their theory of *Pragmatic Reasoning Schemas*. Rather than insight into domain independent syntactic rules of inference or specific memory traces facilitating performance, domain specific reasoning schemas are hypothesized to mediate inference. The appropriate domain or context will invoke a schema which facilitates the identification of the implicit relation asserted to exist between antecedent and consequent. They tested the theory relative to a *permission* schema, but other *causal*, *enablement*, *co-occurrence* schemas etc. are hypothesized to co-exist in the cognitive system. A context was provided for subjects wherein they were to imagine they were customs officials using the rule:

(5.15) If a passenger's form says "ENTERING" on one side, then the other side must include "cholera".

Cheng & Holyoak (1985) ran this version in no-rationale and rationale conditions. In the rationale version they were given a *reason* for why they had to check the forms which could elicit search for counter-examples. In the same experiment a version of the postal rule was used. The subjects were divided into two groups one from Hong Kong, where a similar postal rule was in force, the other Michigan University students. Overall they observed very high rates of falsificatory responding in all conditions. However, the rationale significantly improved performance for all conditions except the postal rule for the Hong Kong group, whose responses remained at the rationale level even when none was provided. They hypothesize that this is due to memory cueing serving to elicit the appropriate rationale from memory. However, the crucial observation was that in rationale versions performance was at the same levels as the specific memory trace condition. This argues for domain specific reasoning processes which are not mediated by *specific* prior experience.

Cheng & Holyoak (1985) also carried out an abstract version of the task, where a permission rationale was provided. This produced a significant facilitation over a no rationale abstract version, where the rule was corrected for syntactic form, both rules included a modal *must*, and explicit negations were included on the cards. Facilitation was observed, which they surmise can only be attributed to the elicitation of the appropriate permission schema. Pragmatic reasoning schema theory is closely allied to the view of inference developed in situation theory. Constraints encode the specific relations which implicitly hold between antecedent and consequent. Moreover, the explicit provision of contexts of interpretation and rationales is consonant with the view that interpretation always proceeds with respect to a context, either explicit in the environmental circumstances of inference and comprehension or implicitly recruited from memory. It could be hypothesized that Pragmatic reasoning schemas provides precisely the data structures required to implement pragmatic context theory. However, there are points of divergence which will lead to the identification of a possible response due to Manktelow & Over (personal communication, but cf. Manktelow & Over, forthcoming).

Interpretation always proceeds with respect to a context (Bransford & Johnson, 1972, 1973; Bransford, Barclay & Franks, 1972; Bransford & McCarrell, 1975). If no context is explicitly provided then people assume one, they recruit appropriate material form memory of *similar*, or otherwise related events. If they cannot, then a text may appear meaningless (Bransford & Johnson, 1972, 1973), or subjects may mis-parse the text (Crain & Steedman, 1985). The computational mechanisms by which people recruit only *relevant* experience from memory to provide contexts of interpretation is one of the most pressing problems of contemporary Cognitive Science. But that people do this is beyond question. What also

seems beyond question is that recruitment is mediated by similarity and analogical processes. If this is the case then there may be no need to invoke stored domain specific, but abstract reasoning schemas. An appropriate "schema" could be generated *on the fly* as a summation of the memory traces which a stimulus evokes either directly or by analogy (cf. Rumelhart, Smolensky, McClelland & Hinton, 1986). Similar processes may mediate the contextual effects hypothesized by pragmatic context theory without the need to explicitly invoke stored data structures like pragmatic reasoning schema.

Moreover, the form of the schema Cheng & Holyoak (1985) provide invites another interpretation. The schema is hypothesized to consist of the following four production rules:

Rule 1: If the action is to be taken, then the precondition must be satisfied.
Rule 2: If the action is not to be taken, then the precondition need not be satisfied.
Rule 3: If the precondition is satisfied, then the action may be taken.
Rule 4: If the precondition is not satisfied, then the action must not be taken.

The rules mirror the truth conditions for the material conditional. However, what meaning is to be attached to the modal terms in the productions? As things stand, their explanatory power is wholly parasitic on the readers pre-theoretic understanding of modal terminology. Before this production set could be treated as a computational theory, the terms must be provided with a semantics, either a denotational semantics, ie. truth conditions are provided, or a procedural semantics, ie. in terms of the causal role of these symbols in the overall production system architecture.

A denotational semantics is perhaps ruled out by the claim that since the productions include modal terminology, and standard logic does not, there can be no mental logic. But this argument succeeds only on the assumption that there are no modal logics, which is false (Hughes & Cresswell, 1968; Stalnaker, 1968; Lewis, 1973). The specific modalities involved are *deontic*, and again systems of deontic logic exist (Castaneda, 1975). There is no reason to assume that Cheng and Holyoak's permission schema could not be formalised as perhaps the non-logical axioms of a modal theory of a permission relation. Alternatively they may well function as the logical axioms governing the behaviour of a deontic conditional (Jackson, 1985, Manktelow & Over, personal communication). The theoretical issues raised here will be dealt with more thoroughly in the conclusions. But it is precisely because Cheng & Holyoak's explicit theory does not rule out this possibility that pragmatic reasoning schema, as articulated, seem inappropriate as the data structures in which to implement pragmatic context theory.

In summary, the conclusions derivable from the observation of the thematic facilitation effect would appear to have turned full circle. The rules which seem to elicit most falsificatory responses relate to deontic contexts where the rule expresses a conventional regulation of some form. As Manktelow & Over (personal communication) observe this implies that people may possess a mental deontic logic. However, there is one notable exception, ie. the original rule used by Wason & Shapiro (1971). Although there have been failures to replicate this result, a pure deontic reasoning account based on deontic logic or reasoning schemas seems, *prima facie*, unable to explain what population differences could account for this anomaly. Moreover, the results obtained using abstract rules and materials in the COST are wholly beyond the scope of such a theory (Johnson-Laird & Wason, 1970; Wason & Green, 1984; Beattie & Baron, 1988).

5.4.1 Thematic facilitation and pragmatic context

The conclusion which work on the thematic facilitation effect appeared to license was that peoples reasoning is content dependent and hence subjects were irrational. The latter inference will not be discussed until the conclusions. However, all the preceding chapters are predicated on the assumption that it is rational to adapt ones strategies to the goal of successful action in the world, not to blindly follow the dictates of some formal inference regime. With this in mind, this section will present an interpretation of these results within the framework of pragmatic context theory.

Prima facie the accounts offered for this effect fail to address two further issues. No relation is admitted between the observations made concerning subjects truth table evaluations (Johnson-Laird & Tagart, 1969; Evans, 1983) and constructions (Evans, 1972) and their selection task performance. The observation of defective truth tables should connect to the theories which purport to describes how the conditional mediates inference. Second, there has been no discussion of the gap that exists between knowing the truth conditions of the conditional and knowing how to fixate a conditional belief. In the strategies outlined at the beginning of this chapter, the combinations which make the rule true or false remained the same, *but* dependent on varying contextual assumptions the strategies appropriate were very different. The observations made with abstract material have served to identify fundamentally different but contextually appropriate strategies for fixating conditional beliefs. However, the thematic facilitation effect appears most efficacious only in contexts which drastically alter the nature of the task: in these versions it is no longer an inductive task.

This tends to render the notion of "facilitation" redundant. Facilitation of a particular response in one task can't be demonstrated by the observation of that response in a different task. In the deontic tasks, subjects no longer have to determine the truth or falsity of the rule but rather on the assumption it is in force whether it has been violated (Yachanin & Tweney, 1982; Pollard & Evans, 1987). However, as Pollard & Evans (1987) observe, this is not simply a function of being instructed to look for violating cases. Similar instructions for the abstract case does not yield facilitation (Griggs & Cox, 1982). There are several observations that need to be made concerning the determinants of subjects behaviour, they relate to:

(1) The provision of thematic content.
(2) The provision of an appropriate context.
(3) The provision of a rationale.

Cheng & Holyoak not only provided subjects with an appropriate context, within that context they provided an appropriate rationale, ie. a reason why subjects should make the check. Although this facilitated performance over the no rationale groups, it was the icing on the cake, so to speak. Subjects were already performing at around 60% correct, even without a rationale which however, raised performance to around 90%. By most metrics used in these experiments Cheng & Holyoak were already observing facilitation. So a rationale provides additional reasons to look for violations. What of the remaining 60% facilitation?

Pollard & Evans (1987) have addressed this question. A two way factorial design incorporated abstract and thematic material with a scenario (context) and without. They discovered a main effect for scenario but not for presence or absence of thematic content, and a significant interaction such that thematic materials plus a scenario produced the best performance. This seems to imply that both (1) and (2) are required to elicit the appropriate falsificatory response. Very little facilitation was observed for the abstract material with a scenario over the condition where it was absent. Their results seem to conflict with Cheng & Holyoak's finding that with the provision of an appropriate, but still abstractly stated, permission rationale for the abstract variant, facilitation was observed. However, there were differences. Cheng & Holyoak used modals in the rules, and the scenario quite repetitively emphasised that the people *must obey* the rule, and that fulfilling "P" is a *prerequisite* for performing action "A". The word "regulation" occurs three times in the instructions. It seems hard to imagine that subjects would not be able to generate sets of relevant experiences given this amount of prompting. In general, these tasks appear to indicate that the effect principally requires the right context or scenario. If this is prompted sufficiently it may well facilitate performance even on abstract versions.

The change in the task involves a switch away from attempting to determine whether the rule is true in an open or closed system to using the rule to guide a certain kind of action. The actions do not involve make eductions concerning what happens next in the world. Rather they involve subjects acting as an arm of the institutions which enforce conventional laws. The scenarios or contexts ensure subjects imagine themselves as performing appropriate actions, however one does not act unless one has an appropriate reason, ie. a goal. Providing the rationale provides subjects with the appropriate goal or reasons for performing these actions. Moreover, they are going to want the rule in the form which is most conducive to performing the appropriate actions.

Conventional constraints are often relatively ubiquitous. Although, violators exist, within their jurisdiction, conventional constraints are supposed to hold without exception. This moderates the effect on inference of possible mitigating background conditions. For example, if Johnny is drinking under age he is still violating the law even if its his birthday, he didn't know the law: he thought he was in another state (even lack of attunement doesn't get you off the hook) etc. According to pragmatic context theory, these manipulations will have had an effect on subjects rule interpretations, the question is what effect?

Conventional constraints are either of the form PA (if precondition, then action) or AP (if action, then penalty). All the rules mentioned are PA rules. The order PA is not the order in which the rules in these tasks are presented, they all have the following form:

(5.21) If ACTION, then must PRECONDITION.

However, once subjects adopt the point of view suggested by the scenario, they need the rule in the appropriate form. For example, checking for under age drinkers is not achieved exhaustively by checking all beer drinkers to see whether they are over 18. Rather, potential violators who look like they are under 18 are checked to see whether they are drinking beer. This reflects the appropriate order of property determination (cf. 5.2.2): subjects require the rule to be interpreted in an action orienting manner. If the rules surface form is not appropriately action oriented, then they may convert to a contextually apposite equivalent which is. What appropriate equivalents does situation theory allow for deontic, conventional contexts?

In situation theory the locus of modality is given by the constraints (cf. 3.3). Modal terminology usually signifies that some such relation is in operation. For example, (I am indebted to Ken Manktelow & David Over for this example)

(5.22) You ought to wear gloves. (said to a nurse who is cleaning up blood)

(5.22) illustrates two features of modals. First, the modals relates two contingencies, one elliptically present in the context, as a conditional, ie. if you clean up blood you ought wear gloves *because* contaminated blood causes AIDS so wearing gloves can *prevent* catching AIDS. Similarly for Cheng & Holyoak's rationale, if you enter the country you must be inoculated against cholera, to *prevent* the native population catching the disease.

Cheng & Holyoaks's rule could be stated in the form of their Rule 4.

(5.24) Not being inoculated against cholera PRECLUDES entry to this country.

All the deontic conditionals used express non-taxonomic constraints. By way of exampling the kinds of inferences licensed by (5.24) let us briefly return to the my hall light example. The background types bear the same relation as expressed in (5.24) to the indicated type, eg. the electricity not being on (E = 0) precludes the light coming on (L = 1), ie. they are negatively causally related:

(5.25) $E = O \ =| \ L = 1$

Equally, the light being on precludes the the electricity being off:

(5.26) $L = 1 \ =| \ E = O \ <-> \ L = 1 \ => \ E = 1$

and this is equivalent to the constraint that the light being on involves the electricity being on. However, observe that the electricity being on (E = 1) does not involve or preclude the light being on (L = 1), nor does the light being off involve or preclude the electricity being on.

The following set of propositions describe the situation with my hall light and the electricity supply, taking T: *<<Electricity-on; 1>>* and T′: *<<Light-on; 1>>*

Rule 1′: $s \models$ <<[n]=>, T′, T; *1*>>

Rule 2′: $s \not\models$ <<[n]=>, ¬T′, T; *1*>> & $s \not\models$ <<[n]=|, ¬T′, T; *1*>>

Rule 3′: $s \not\models$ <<[n]=>, T, T′; *1*>> & $s \not\models$ <<[n]=|, T, T′; *1*>>

Rule 4′: $s \models$ <<[n]=|, ¬T, T′; *1*>>

These propositions exactly mirror Cheng and Holyoak's (1985) pragmatic reasoning schema. However, they are not theoretical constructs but a characterisation of the content of peoples mental states. This characterisation is fitting since being inoculated is just one of

the many conditions you must satisfy to enter the country, others being, eg. that you want too because you have to go to a conference etc. The modal terminology has been systematically replaced by the relations of involvement and preclusion. And the contextual grounds for rules 2′ and 3′, are made explicit. In the context of my hall light, the electricity being on does not involve my hall light being on: the switch may not be turned, and since the electricity being on is what enables the turning of the switch to turn on the light, the electricity being on cannot preclude the light coming on. Similarly the precondition being satisfied does not involve taking the action, and it does not preclude it either. Observe that rule 4′ is in the only rule supported in the situation whose discovery order matches the order in which I suggested subjects would need orient their rule interpretation to guide their actions once they have adopted the point of view suggested in the scenario.

Generally, it is not the case that if T involves T′, then ¬T′ precludes T. Rules 1′ and 4′ are not *always* or indeed usually equivalents. Recall the original my hall light example, switch A being turned involves the hall light going on, but the hall light not being on does not preclude switch A being turned, the bulb may be broken. It is only relative to other pragmatic factors concerning the ubiquity of the conventional constraints when in force (or the enabling relation between the electricity being on and the lights going on), which licenses 4′ given an expression of 1′.

To guide their actions appropriately subjects adopt the interpretation in 4′. When making inferences subjects need not have a whole pragmatic reasoning schema available, rather than this one interpretation. One highly resilient observation in selection task performance is the almost universal selection of the *p* card, in many studies the *p* card only as a selection is the second most frequent. If subjects were interpreting the rule as in 4′, to guide their actions, and subsequently reasoning in the converse direction to select the *p* card then a reversal of this common result would be expected. The negation of the consequent of the surface form of the rule is now the antecedent in 4′, hence rather than the observation of *p* card only selections, ¬*q* card only selections would be expected. However, if a whole pragmatic reasoning schema were available, then subjects could use the productions in 1 and 4 (above), and no switch would be expected. However, although Cheng & Holyoak do not report their results in sufficient detail to check this prediction, Pollard & Evans (1987) do. The prediction of some ¬*q* card only selections was supported in their thematic/scenario version, where 21% of subjects selected the ¬*q* card only, while no *p* card only selections were made. 71% of subjects made the *p* and ¬*q* card selections. If subjects do make the conversion to the interpretation in 4′, then modulo that interpretation subjects responses accord with *verification*. That is, the eductions licensed by 4′ form a mirror image of

standard abstract selection task performance, as indeed do Pollard and Evans thematic/scenario condition data. So once the appropriate action oriented interpretation is adopted, subjects again are using the rules to make information gaining eductions.

The anomaly of the original Wason & Shapiro (1971) result can be interpreted quite readily in terms of the rule expressing a relational taxonomic constraint. Since this is the case, a unified interpretation exists for the cards which could cue the negative confirmation strategy outlined for the taxonomic schema presented in section 5.2.2. The failure to replicate observed by Manktelow & Evans (1979), can only be put down to population differences. Perhaps Manktelow & Evans (1979) subjects made an open system assumption (cf. 5.2.2). This account clearly does not constitute an *explanation* of the replicative failure but it does provide a plausible account of the differences in the content of subjects' mental states, between experiments, which could characterise the differences in subjects' obsvered behaviour.

Summary

This chapter began with a detailed account of the strategies licensed in Wason's selection task by a non-taxonomic and a taxonomic interpretation of a conditional. It was then shown how the disjoint expression of the rule in the selection task function to elicit the standard predictive cycle strategy which warrants eductive inferences which accord with the strategy of the verification. It was then shown how the account generalises to the therapy experiment of Wason (1969). The thematic "facilitation" results were then reviewed and it was argued that the manipulations employed radically changed the nature of the task. However, once the appropriate action orienting interpretation is adopted it was shown that subject's eductive behaviour proceeded as predicted by pragmatic context theory. The exceptional case of the Wason & Shapiro (1971) result was also shown to have a rational explication in terms of a taxonomic interpretation of the rule employed.

In the next chapter the data obtained form Evans' negations paradigm in the selection and truth table table tasks is reviewed. It is argued that the conclusion that irrelevant responding is attributable to either an attentional matching strategy or shallow linguistic processes is unwarranted. This discussion will provide the basis for experiments to be reported in chapter 7.

Chapter 6: The Evans Negations Paradigm

6.1 Introduction

In the preceding chapters the crucial feature of pragmatic context theory has been the demonstration that general considerations of context dependence motivate the concept of partial interpretation. In chapter 4, the consequences of partial interpretation for the strategy of falsification were outlined. It was argued that if this strategy is motivated either by appeal to *modus tollens* or the semantics of the universal quantifier, partial interpretation still rendered falsification an unsound strategy of confirmation. In chapter 5, it was argued that the interpretative process which incorporates varying contextual assumptions either given in the environmental circumstances or from prior beliefs, were preconscious and automatic, in line with Wason & Evans (1975) dual process theory.

It was mentioned that Evans has a different interpretation of these *Type 1* processes. In experiments, which will be reviewed below, Evans systematically varied negations in the conditional and discovered that subjects appeared to ignore the negations and simply *match* named values. However, in protocol data (Wason, 1969; Wason & Evans, 1975) subjects were observed (i) to realise that a true antecedent and false consequent instance made a conditional false, and (ii) gave *post hoc* logically sound justifications for selections. This anomaly provided the principle motivation for the dual process theory. However, it also gains credence from distinctions between preconscious and conscious processes in the perceptual psychology literature (eg. Neisser, 1967). *Matching bias* was the result of selective attention processes (Evans, 1983a). One discrepancy between Evans' (1983a) view of preconscious processes and earlier conceptions concerns the difference between *mental set* (Duncker, 1945, Luchins & Luchins, 1950) and selective attention. Neisser (1967) identifies preconscious processes as pre-attentive. This is based on perceptual studies involving rapid holistic processes which occur automatically prior to selective attention picking out features for further processing. Evans (1983a) argument reverses this process: preconscious processing is determined by a *mental set* which directs attention to named values. The view that will be adopted here (cf. conclusions) is that the original distinction, whereby preconscious processes are pre-attentive and holistic better captures the early interpretative process where all sources of information enter into the determination of an interpretation prior to higher level Type II processing.

Issues of process to one side, the principle reason for focusing on Evans negations paradigm involves the interpretation of this data as implicating an apparently non-logical matching bias as determining subjects irrelevant responses in truth table tasks. It has already been argued that the occurrence of irrelevants is a function of partial interpretation modulo the context dependence of the early interpretative process. Hence, there is a clear disagreement concerning the nature of Type 1 processes. Although pragmatic context theory concurs with the view that these processes are non-logical insofar as they are not determined by some standard logical inference regime, this does not mean they are not constitutive of the interpretative process. This implies that interpretative factors should be the principle cause of subjects behaviour even when negations are systematically varied. The interpretations licensed by pragmatic context theory will be non-standard, so the data on Evans negations paradigm stands in need of re-evaluation in the light of these interpretations. Of the many predictions derivable from pragmatic context theory, this was considered the most important to validate empirically since Evans' interpretation of these results casts doubt on the theory's central concept: partial interpretation.

The structure of this chapter is as follows. In the next section the data on Evans negations paradigm will be critically reviewed and assessed. This will be followed by an account of the situation theoretic interpretations of conditionals containing negatives. In the last section the experiments which will be reported in the next chapter will be introduced, and the motivation provided for the procedures and materials employed.

6.2 Critical review of work using Evans negations paradigm

6.2.1 Truth table construction and evaluation tasks

Some terminology first needs to be introduced. The systematic variation of negatives in conditional rules means that the combinations of p and q and their respective negates $\neg p$ and $\neg q$ shown in Table 1 correspond to the True/True (TT), True/False (TF), False/True (FT) and False/False (FF) combinations in the standard truth table for the material conditional. Most of the discussion of these tasks goes on in terms of these logical cases, rather than card cases. This already presupposes a recurring assumption made in analysing this data: there is no semantic interaction between negation and the conditional. Evans (1972) conducted a truth table construction task in which the first observations of matching bias were made. Subjects were presented with a 4 × 4 array of cards containing various

Table 6.1

Combinations corresponding to standard Truth-table cases for Rules with Negated constituents.

Rule	TT	TF	FT	FF
(1) If p, then q	pq	$p\neg q$	$\neg pq$	$\neg p\neg q$
(2) If p, then ¬q	$p\neg q$	pq	$\neg p\neg q$	$\neg pq$
(3) If ¬p, then q	$\neg pq$	$\neg p\neg q$	pq	$p\neg q$
(4) If ¬p, then ¬q	$\neg p\neg q$	$\neg pq$	$p\neg q$	pq

coloured shapes (circle, triangle, cross, square; red, yellow, green, blue). The four rule variants were presented one at a time and the subjects' task was too pick two of the cards from the array and place them side by side such that they made the rule true or false dependent on whether they had been instructed to verify or falsify it respectively.

Two principle motivations for this experiment are cited by Evans (1972). First, Johnson-Laird and Tagart (1969) had claimed support for a defective truth table using an evaluation task. Evans (1972) observed that the evaluation procedure involves providing subjects with an "irrelevant" category for classifying whether instances falsified or verified a conditional rule. This invited the criticism that subjects may well have used the category simply because they felt they ought to. To avoid this Evans (1972) developed the construction task so that the irrelevant category could be inferred from a subjects' failure to construct an instance. Second, previous experiments in which negatives had been introduced preserved the p *implies* q logical relationship. This could allow truth and falsity to be confounded with affirmation and negation. Evans (1972) varied the rules as in table 6.1. He had also observed that inference by *modus tollens* is less often produced in the presence of negated antecedents (Evans, 1972a). It was therefore hypothesized that fewer TF cases would be constructed to falsify rules with negative antecedents. He observes (Evans, 1972:194) that, "failure to make modus tollens is logically inconsistent with the belief that TF falsifies".

The results of this experiment will be reported in some detail. Subject's initial response data (they were permitted to make as many constructions as they felt appropriate) is summarised in table 6.2. Only the first verifying and first falsifying construction was included in this table on the assumption that they, "would be the psychologically "strongest", on the grounds they occurred most immediately to the subject" (Evans 1972:195). The prediction that fewer TF cases would be constructed to falsify rules with negative antecedents received confirmation (p = 0.0005, sign test) and there were significantly more TF cases constructed to falsify rules with negative consequents (p = 0.012, sign test). Evans (1972:196) takes

Table 6.2

The Frequency with which subjects gave each truth table case on their initial verification and initial falsification of each rule. (n = 24)

Rule	Verification				Falsification			
	TT	TF	FT	FF	TT	TF	FT	FF
(1) If p, then q	24	0	0	0	0	17	0	6
(2) If p, then not-q	24	0	0	0	0	23	0	0
(3) If not-p, then q	24	0	0	0	0	7	15	2
(4) If not-p, then not-q	22	0	0	2	0	13	2	8

these results as evidence that, "subjects prefer to match rather than alter the values named in the rules, which we shall refer to as matching bias".

Is it legitimate to take only subjects first falsifying or verifying constructions into account in establishing this conclusion? In the discussion Evans observes that the overall results for the *if* $\neg p$, *then* q rule indicated that subjects were treating it as exclusive-OR. Exclusive-OR is equivalent to the denial of material equivalence, hence there are two potentially falsifying values TF and FT. Both will figure in subjects responses. That one or the other is initially preferred represents a *bias* which may be explained either by attentional *or* interpretative *processes*. If it is shown that subjects exclusive-OR interpretation was appropriate, then this tends to invalidate the matching bias conclusion. However, the bias may also be observed in the total response data.

Table 6.3 shows subjects total response data, which was not reported in this form in Evans (1972), but it could be extracted from another table in which Evans constructed psychological truth tables form the total response data (cf. next chapter). To determine whether matching bias is observed in the total responses sign tests could not be used as the data was not

Table 6.3.

The total frequencies with which subjects gave each truth table case on verifying and falsifying each rule. (n = 24)

Rule	Verification				Falsification			
	TT	TF	FT	FF	TT	TF	FT	FF
(1) If p, then q	24	0	2	4	0	21	7	8
(2) If p, then $\neg$q	24	0	5	10	0	23	1	1
(3) If $\neg$p, then q	24	1	2	11	0	15	18	4
(4) If $\neg$p, then $\neg$q	22	2	4	7	0	18	7	9

reported by subject. However, at least an indication can be gained using the albeit inappropriate χ^2 test. The comparisons are falsification TF, (1) & (2) against (3) & (4), the difference was not significant ($\chi^2 = 0.83$, $p > 0.20$), and falsification TF, (1) & (3) against (2) and (4), the difference was not significant ($\chi^2 = 0.21$, $p > 0.20$). There would appear to be little evidence for matching in Evans (1972) total response data. Evans (1972) does conclude that the data provides support for the defective truth table account, generally irrelevant responses could be inferred for false antecedent instances.

On the assumption that the exclusive-OR interpretation was in some sense appropriate what requires explaining is why subjects initially preferred FT when falsifying the *if* $\neg p$, *then* q rule since both pq (FT), and $\neg p \neg q$ (TF) falsify it. The twin observations of (i) subjects apparent XOR interpretation, which (ii) was the artefactual cause of the significant matching result in the initial response data, represent the primary motivation for subsequent replication of this task to be reported in the next chapter.

Evans (1982) returns to the original Evans (1972) data to motivate the claim that the occurrence of irrelevants on that experiment were actually a function of matching bias. The argument proposes that the standard logic of the material conditional be retained in the interpretational or logical component, with deviations being explained by non-logical biases, in a response-bias component of the cognitive system. Evans presents the data pooled over the four rules, by logical case, and by "matching" case, ie. card cases in the order pq (0 mismatches), $p\neg q$ & $\neg pq$ (1 mismatch), and $\neg p \neg q$ (2 mismatches), as in table 6.4. Evans (1982:140-142) observes that the modal responses (in italics) by logical case were clearly supportive of a "defective" truth table account (Wason 1966). However, looking in the "irrelevant" column for the data arranged by matching case reveals a trend towards these non-constructed choices as the number of mismatches increases, ie. subjects are more likely to construct instances whose items match those named in the rule. On this basis Evans (1982) adduces two alternative hypotheses to explain the high occurrence of "irrelevants" on the FT and FF cases:

(i) People possess a defective truth table in the logical component.

(ii) People *weight* the logical component less in these cases and therefore "irrelevants" arise from matching bias.

Evans (1982) plumps for (ii) on the basis of the following considerations, "If we assume that the matching effect is really suppression of responding on mismatching cases, then we can resolve this problem. Is there still a higher irrelevant rate on false antecedent cases if

Table 6.4

Evans' (1972) results pooled over the four rules and classified by (a) Logical Case and (b) by Matching Case. Results are percentage frequencies (N = 24).

Case	Classification		
	True	False	Irrelevant
(a) Logical			
TT	*99*	0	1
TF	3	*80*	17
FT	14	34	*52*
FF	33	23	*44*
(b) Matching			
pq	34	52	14
p¬q	41	33	26
¬pq	40	27	33
¬p¬q	34	25	41

Table 6.5

The percentage frequency of construction of each logical case on rules where they constitute a double match, pq, for Evans (1972) data. *(N = 24).*

Logical Case	Rule	% Frequency		
		True	False	Irrelevant
TT	If p, then q	100	0	0
TF	If p, then not-q	0	96	4
FT	If not-p, then q	8	75	17
FF	If not-p, then not-q	29	38	33

we look only at rules where such cases match" (*p*141). Table 6.5 is the table produced in Evans (1982) based on the Evans (1972) data.

Evans concludes, that "Although there is more "irrelevant" responding to FT and FF, it is relatively low in cases that match. Thus the evidence for the defective truth table account may be a partial artifact of greater susceptibility to response bias on these cases" (Evans 1982:141). However, this argument is not independent of the prior observation concerning subjects apparent XOR interpretation of the *¬p, q* rule form. It was observed above that this interpretation could be primarily responsible for the significance of the matching predictions in subjects initial response data. It just so happens that when restricting the data to the double match cases, the fact that FT was chosen most frequently as falsifying for this rule shows up as a suppression of irrelevant responding. However, it could equally be a

function of interpretational changes, as mentioned above. Moreover, although it is not clear cut in Evans data, a similar argument may account for the apparent suppression of irrelevant responding for the *if* $\neg p, \neg q$ rule form.

However, Evans argues that if there were an interaction between negation and the conditional along the lines being suggested here, another observation is relevant. The TF case (cf. table 6.4(a)) also revealed a significant increase in irrelevant responding as the number of mismatches increased: 0-mismatches: 1 irrelevant; 1-mismatch: 3 & 4; 2-mismatches: 8. Evans does countenance the possibility that this is due to the difficulty of processing negation (Wason, 1959, 1961, 1980). He observes, however, that matching seems to primarily affect false antecedent instances, ie. FT and FF, which argues against a general processing deficit. There is some inconsistency in arguing against a negations-conditional interaction on the grounds that irrelevants also increase for TF instances as a result of matching, while arguing that a negative processing deficit explanation won't do because the effect is only found for FT and FF instances.

In chapter 2 it was observed that negation in situation theory functions to identify well defined contrast classes relative to dimensions of variation in which a negated constituent participates. If it is a binary dimension, ie. an antonym eg. hot/cold, then the negation identifies the antonym of the negated constituent. The antonymic relation can be contextually defined, eg. in a set consisting of only Russians and Germans, ¬German identifies Russian. Similarly, a whole dimension of variation may be defined, eg. ¬tea, identifies a contrast class consisting of drinks (the superordinate category), but in a context where tea is being drunk, coffee is the most felicitous contextually defined opposite. However, if a negated constituent does not fall into some natural dimension of variation or taxonomy, then the contrast class is less determinate. In these cases the human interpreter may be unable to use his stored beliefs but rather the environmental circumstances in the attempt to pick out the relevant contrast class. In accordance with pragmatic context theory, this would be expected to affect interpretation. In Evans (1972), the card array could provide the circumstantial means by which subjects attempt to determine the relevant contrast class. This view of negation is inherently *constructive*, given the negated constituent and prior beliefs or the environmental circumstances, the human interpreter must actively construct the contrast class. The processes involved may well lead to characteristic errors and *biases* especially when the contrast class is ill defined. Again the processes involved could be responsible for the observed matching phenomena. The experiments to be described in the next chapter attempt to distinguish between a negative-conditional semantic interaction account, a constructive negation account and matching. Precise hypotheses are formulated in the last

section of this chapter.

Evans (1983b) hypothesizes a possible linguistic source for matching bias. Wason (1965) observes that the normal pragmatic function of negation is to deny presuppositions. For example, "There are sea dwellers that are not fish" seems to be a pragmatically felicitous utterance when denying a prior assertion to the effect that "All sea dwellers are fish". Pragmatically the *topic* of conversation is still fish. Evans goes on to argue that if explicit negations were incorporated in the instances in a truth table task, then subjects may not discard an instance as irrelevant because it now shares the same topic as the rule. Since mismatching instances are just those where an instance fails to match the named value, and it was observed (cf. above) that there appears to be a trend for more irrelevants in these cases, this should lead to a reduction in irrelevant responding. It is then argued that any interpretational account could have nothing to say concerning this procedural change, since it would rely on suggesting that the matching phenomenon is a function of the effect of negatives on initial *rule* interpretations. The argument to this conclusion relies on separating two levels of processing one linguistic or *heuristic* and the other interpretational, semantic or analytic. It is assumed that subjects make an initial pre-interpretative parse of the rule. The "linguistic" information gleaned then interacts with the instance to determine whether further semantic processing occurs to determine how the instance bears on the rules truth or falsity. Two "linguistic" heuristics determine subjects irrelevant responding. First, the negative topic function described above. Second, a further "linguistic" heuristic, provided by the phenomenon of the irrelevance of false antecedent instances. The latter will affect both groups, ie. a group of subjects who perform a task containing explicit negations on the instances and those that do not.

The task employed was similar to the construction task (1972) but instead of constructing instances subjects were presented with all 4 possible instances for each rule and asked to make a *conform*, *contradict* or *irrelevant* evaluation. The prediction was derived that there would be less matching for an *Explicit* (E) group than for an *Implicit* (I) group (the latter having members of the contrast class, ie. other numbers or letters, on the instances). This prediction was confirmed using an antecedent (AMI) and consequent matching index (CMI):

(AMI) The frequency of irrelevant responses on $\neg pq$ and $\neg p\neg q$ instances minus those on pq and $p\neg q$ instances.

(CMI) The frequency of irrelevant responses on $p\neg q$ and $\neg p\neg q$ instances minus those on pq and $\neg pq$ instances.

Since each subject performed the task twice there was a potential for 2 irrelevant responses per instance per rule form. Hence, both indices range between +16 and -16, a positive value indicates matching. Both were significantly larger for the I-group than the E-group. Moreover, computing a *Logic* index permitted comparison between groups for consistency with the logical response. This was restricted to affirmative antecedents. The index was significantly higher for the E-group. Evans (1983b:642) argues that:

(6.3) "These findings are irreconcilable with the view that matching bias is simply a result of the effect of negatives on the interpretation or initial representation of the conditional rule itself. It is not a change in the rules but in the form of the instances which is making the difference. Furthermore, the logical information conveyed in the instance is unchanged; it is the *linguistic expression* of that information which is critical".

However, in the discussion it is allowed that if matching is the result of a linguistic topic function, then all matching should disappear for the E-group. It is suggested that there may well be some residual processing deficit incurred due to the presence of negations on the instances, in accordance with Wason (1969, 1961, 1980). Moreover, since the manipulation is only supposed to affect the early heuristic stage the ratio of true false judgments should remain the same between conditions, which was not observed especially for FT and FF instances. It is therefore concluded that the instances may be affecting the analytic processes directly in some cases. The following critique of this experiment is going to be quite detailed since, *prima facie*, the results seem to argue strongly against the position being developed here.

(1) The distinction between an essentially linguistic or heuristic level of processing and a deeper semantic or analytic level may be spurious modulo the processes attached to the heuristic level. First, the concept of topic is a discourse function (Haiman, 1978; Akatsuka, 1986). *Prima facie* isolated rules expressed by means of a conditional would appear devoid of an appropriate discourse. The explicit mention of an item appears to be conflated with the item being the topic of the sentence. However, it is debatable whether the concept of *topic* is applicable to individual sentences. Moreover, in the linguistic literature (Haiman, 1978), it is not the case that both antecedent and consequent of a conditional are topic markers, rather this is a function identified with the antecedent. The antecedent is hypothesized to introduce given or shared knowledge (Clark, 1977) into a discourse (Haiman, 1978). These functions are also identified in the psychological literature as being crucial to *understanding discourse*, ie. the are implicated in accessing the stored knowledge required for comprehension (Clark, 1977). This is hardly a a shallow heuristic level of processing, but rather about as deep as one could get, ie. after initial syntactic and semantic

processing, these pragmatic functions of linguistic expressions aid in integrating the interpretation with appropriate world knowledge.

However, in part this could be a terminological quibble. *Topic* has a certain technical application which perhaps carries unintended connotations. The more neutral concept of *aboutness* could be appealed to, in some sense a negated constituent is still *about* that constituent. However, this is contextually bounded. Take for example:

(6.4) If its acid, then it turns the litmus paper blue.

(6.5) If its acid, then it does **not** turn litmus paper blue (...it turns it red).

(6.5') No, if its **not** acid, then it turns litmus paper blue (...alkalies do that).

Negations are used to make denials, which *presupposes* an assertion to deny. In (6.4) an assertion is made which (6.5) and (6.5') deny. In (6.5) interest *focuses* on what colour acid turn litmus paper, in (6.5') it focuses on what substance turns litmus paper red. In this example, the negation functions to focus attention on a highly constrained contrast class. In (6.5) the consequent is *about* what colour acid turns litmus paper which is red, in (6.5') the antecedent is *about* the substance which turns litmus paper blue which is alkalies. In this constrained binary situation the negation identifies determinate members of a contrast class. However, in other contexts a negation may not focus attention on an item of a well defined contrast class, Evans employed the following example:

(6.6) I did not go for walk.

If I suddenly disappeared from my terminal and on my return a colleague says, "Did you enjoy your walk", I may utter (6.6) in reply but not thereby say anything about what I did during my absence. The information conveyed is simply the non-occurrence of my having been for walk. In this case, (6.6) is still *about* going for a walk. This distinction has already been introduced in chapter 2. The sameness of topic thesis seems only to apply to cases where there is no well defined contrast class. But in the task the subjects know that there is a letter on one side of each card and a number on the other.

The fact that the denial *presupposes* a prior assertion, raises an issue which would seem to run together the two heuristic processes. What if I said, "the King of France is bald", and you replied "the king of France is not bald", have you successfully denied my assertion?. Well it appears that since there is no King of France, I have said nothing with a determinate truth value, and hence your denial is equally vacuous. My assertion and your subsequent denial rely on the existential presupposition that there is a King of France. Just as

using a negation to make a denial presupposes a prior assertion; an assertion appears to presuppose the existence of the objects about which it is made. Note that in Evans' example, in my reply to my colleague, I have precisely denied the presupposition upon which his question was predicated, ie. that there was a walk to enjoy. Presuppositional phenomena constitute one of the primary motivations for introducing truth value gaps, a third truth value (?) or partial interpretation. Moreover, the debate concerning whether this is a semantic phenomenon has raged since Frege (1892, cf. Russell, 1905, Strawson, 1950). If a third truth value or partial interpretation is allowed then the first candidate for re-interpretation is the conditional: a conditional only makes an assertion given the truth of its antecedent in which case the assertion made is the consequent, otherwise it does not possess a determinate truth value.

There has been much debate as to whether presupposition should be dealt with by semantics or pragmatics (Gazdar, 1979; Seuren, 1985). However, the distinction between pragmatics and semantics is only a formal convenience of the linguist, all knowledge sources must interact in the human interpreter. The level at which these phenomena are discussed is semantic/pragmatic, they are crucially implicated in a our normative conceptions of meaning and interpretation, and indeed the issues involved are still some of the deepest facing natural language semantics. On pragmatic context theory and any reasonable theory of the interpretative component these phenomena are going to be central. Both the phenomena described are usually attributed to semantic/pragmatic factors. It therefore seems inappropriate to attribute their causes to early processing heuristics rather than perhaps the very deepest level of the interpretative process.

(2) The kind of theory Evans is arguing against relies on an initial interpretation of the rule subsequently determining how each instance is handled. Pragmatic context theory views the interpretative process as involving all knowledge sources, including prior beliefs *and* circumstantial features of the environment. Negation especially is not tied to a particular representation, the representations it contributes to constructing will depend crucially on context. This is because negation is treated constructively: the processes involved will undoubtedly interact with the context provided by the instances, as is conceded in the discussion section. Moreover, pragmatic context theory hypothesizes that the reason for the processing deficit incurred through the use of negatives is precisely caused by the constructive processes involved. Since it is also conceded that some difficulty may be imposed by the preponderance of negations, this is all to the good: a constructive view of the interpretative processes involved may provide an explanation for the deficit. Generally, rather than suppose that the occurrence of irrelevants is due to two different kinds of heuristic process,

one affected by the implicit/explicit manipulation the other not, it seems that the difference could be one of degree involving the same underlying interpretative process. Furthermore, on each instance each subject had the rule available. The only way to test whether a shallow pre-interpretative parse of the rule was responsible for discarding an instance as irrelevant would be to take reading time/response time measures and correlate these with the occurrence of irrelevants.

(3) In the introduction it is conceded that a great deal of irrelevant responding is due to the perceived irrelevance of false antecedent instances (FT & FF). The issue over whether this phenomenon is due to a linguistic heuristic too one side, he nonetheless fails to correct for the possible effects of this additional factor. Recall matching produces irrelevants as a function of the *aboutness* heuristic *not* the irrelevance of false antecedent instances. However, the two *matching* indices are computed for the *whole* response record including the irrelevants which are not hypothesized to be a function of matching. At least with no indication to the contrary it must be assumed that this is the case. However, taking the double mismatch case, the percentage of irrelevants observed for false antecedent cases makes up 32.25% of the total percentage of 45%. That is, irrelevants occurring for false antecedents, where they are appropriate for reasons other than matching, are responsible for the observed increase in irrelevants for the double mismatch case.

The numbers of 0, 1, and 2 mismatching/matching instances is exactly symmetrical between true and false antecedent cases. This means that TT and TF instances, match and mismatch *exactly* the same number of times as FT and FF instances. Since, the *aboutness* early processing heuristic is supposed to *block* further processing given a mismatch, but especially on double mismatches, then this process would predict *exactly* the same numbers of irrelevants between true antecedent and false antecedent instances. True/false assessments are only carried at the deeper analytic level: they have to get past the heuristics first. However, in Evans data, for the I-group only 26% of all irrelevants were observed for true antecedent instances as opposed to 74% for the false antecedent instances, the figures were 18% and 82% respectively for the E-group. Since matching makes symmetrical predictions this means that 48% of all irrelevants in the I-group and 64% in the E-group are solely due to the irrelevance of false antecedent instances and processing negatives. However, the defective truth table account would predict no differences between 0, 1 and 2 mismatching cases, which clearly there were, but this is no more than was apparent from Evans (1972).

(4) There are some quite marked population differences between Evans (1983b) and Evans (1972). First, for the TT case in Evans (1972) for the *if*$\neg p$, $\neg q$ rule form 96% of subjects

made the correct verifying construction against 68% who evaluated this instance as true for the I-group in Evans (1983b). Moreover, in Evans (1972) results on the false antecedent cards were broadly in line with the defective truth table account. As indicated above (Table 6.8(a)), the predominant response for false antecedent instances was irrelevant. The majority of these responses were observed for affirmative antecedent rules (this will prove important later on). Moreover, Evans (1982:142) observes that the response profiles observed in Evans (1972) are "remarkably similar" to evaluation tasks (Evans, 1975; Evans & Newstead, 1977). However, in Evans (1983b) for negative antecedent rules *all* the modal responses accorded with the material biconditional, ie. for both the I-group and the E-group, FT was treated as false and FF as true. This only occurred for the *if* $\neg p,q$ rule form in Evans (1972). For Evans (1983b), for the *if* p,q rule form FT was treated predominantly as false in both I and E-groups. The FF instances was treated as irrelevant in the I-group but as true in the E-group. For the *if* $p,\neg q$ rule form FT was treated as irrelevant in both groups and FF was treated as true. Apart from the FT/*if* p, *then* $\neg q$ case, subjects would appear to be treating the rules as biconditionals throughout, but for certain cases this becomes a more obvious interpretation in the E-group. This does not replicate earlier studies.

(5) There may be a far simpler account of why subjects made less irrelevant responses in the E-group. As Evans (1983b) observes, for some cases subjects are confronted with four negations. For the *if* $\neg p$, $\neg q$ rule form, eg. "if not B, then not 4", they may be presented with an instance which has, "The letter is not B and the number is not 4" on it. Since it is already well established that negations produce a processing deficit, it seems entirely plausible that when confronted with this many negations subjects may simply adopt a heuristic strategy, but one that is more perceptual than linguistic. It seems that the population each sample was drawn from has a general tendency to interpret the rule as a biconditional. However, for some false antecedent instances in the I-group this is found harder. But for the E-group a very simple strategy can be adopted to always get the right biconditional interpretation. Pick any rule, say *if* $\neg p$, q, then if the instance either completely matches ($\neg pq$), ie. perceptually including the negation, or completely mismatches ($p\neg q$), then classify as true, if only one component matches ($\neg p\neg q$; pq) classify as false. This of course works for all rules/instances. 76% of subjects total responses in the E-group could be accounted for by this strategy. It would also account for the differences between the groups, since without a simple perceptual matching strategy I-group subjects cannot circumvent the normal interpretative process. They nonetheless predominantly treated the rules as biconditionals, 60% of total responses accorded with this interpretation. The 16% shift could well account for the observed differences.

In conclusion, the motivation for Evans's heuristic level processes appears questionable. The phenomena upon which they are based are generally agreed to be the function of quite deep semantic factors. The serial order of the heuristic-analytic processes involved predicts symmetrical responses between false and affirmative antecedent instances which renders suspect the procedure of computing a *matching* index over the whole response record. Moreover, the population differences observed suggests that the E-group subjects could be adopting a simple perceptual matching strategy. Evans' arguments only succeed against a certain view of the interpretative process which pragmatic context theory rejects. However, the issue of the source of the matching phenomenon can only be resolved via a competence model of interpretation which provides a rational basis for the empirical observations in this domain. The theory Evans objected to clearly fails in this regard. So the question becomes: does pragmatic context theory fare any better? This is an empirical question which will be addressed by experiments in the next chapter.

6.2.2 Selection Tasks

Evans (1972) speculated that matching bias may be present in the selection task. There were good grounds for this assumption. Using only affirmative rules, subjects could simply be matching since both verification and matching bias make the same predictions for this rule. Evans & Lynch (1973), therefore conducted a selection task using all rule variants. Again some terminology needs to be introduced. Rather than talk in terms of card case, logical case is used, ie. true antecedent (TA), false antecedent (FA), true consequent (TC) and false consequent (FC) (cf. table 6.6). Evans & Lynch (1973) argued that to assess the effects of matching logical case should be kept constant. Four predictions were derived which have subsequently been taken as the *sine qua non* of the presence of matching bias. They were as follows:

Table 6.6

Card case to logical case conversion table.

Rule	TA	FA	TC	FC
(1) If p, then q	p	$\neg p$	q	$\neg q$
(2) If p, then ¬q	p	$\neg p$	$\neg q$	q
(3) If ¬p, then q	$\neg p$	p	q	$\neg q$
(4) If ¬p, then ¬q	$\neg p$	p	$\neg q$	q

(i) There will be more TA selections when there is an affirmative antecedent than when there is a negative antecedent.

(ii) There will be more FA selections when there is a negative antecedent than when there is an affirmative antecedent.

(iii) There will be more TC selections when there is an affirmative consequent than when there is a negative consequent.

(iv) There will be more FC selections when there is a negative consequent than when there is an affirmative consequent.

In Evans and Lynch (1973) all these predictions were significantly borne out using one-tailed sign tests. Table 6.7 shows the results obtained. Each prediction is within a column and the figures in italics indicate the selections predicted to be the most frequent on the basis of matching.

In the discussion of the results, Evans and Lynch (1973) claim that these predictions are independent. However, this is only the case when the data is analysed ignoring any differences in rule form. However, it is a frequent observation that negative information is harder to process than affirmative information (Wason, 1959, 1961). This would suggest that unless some correction for the processing deficit incurred as a result of employing negatives is used, these results need not be taken as directly supportive of matching. Unless it is argued that matching behaviour is a consequence of that deficit, an interpretation which Evans & Lynch do not countenance (although, Evans (1983a:141) does argue that the pragmatic function of negations to deny presuppositions (Wason, 1965) may be implicated in matching phenomena, cf. above). Moreover, many of the observations of higher frequencies of selections would have been predicted by verification/falsification if the analyses were carried out within rules.

The appearance of verification is treated as manifest in the results when there are more TA

Table 6.7

The frequencies with which subjects selected each alternative on each rule (n = 24).

Rule	TA	FA	TC	FC
(1) If p, then q	*21*	2	*12*	8
(2) If p, then ¬q	*22*	1	2	*14*
(3) If ¬p, then q	14	*7*	*14*	10
(4) If ¬p, then ¬q	13	*11*	7	*18*

than FA selections (TA > FA), and more TC than FC selections (TC > FC). Falsification is manifest when there are more FC than TC selections (FC > TC). This appears to indicate that verification and falsification make less detailed predictions in the data than matching bias. However, this is simply an artifact of how the data is organised. Whether logical case or card case is kept constant will determine the detail of the predictions made. Moreover, if concern centres on interpretational differences between rule forms, then it makes sense to see what predictions each strategy makes when rule form is kept constant.

In table 6.8, the predictions made by each of matching bias, verification bias and falsification are illustrated, (i) when logical case is kept constant and (ii) when card case is kept constant. By logical case (card case), 'O' indicates where a large number of responses is predicted and '•' a smaller number. Ignoring rule form for the moment, table 6.8 reveals two levels at which these strategies make predictions dependent on whether logical case or card case is kept constant. (There is nothing sacrosanct about keeping *logical* case constant, when the predictions made by each strategy for each rule form are derived it is wholly irrelevant as to which is chosen. The only difference it makes concerns which rule forms are collected together to derive predictions at a less detailed level of analysis.) The two levels separate how many predictions are derivable from each strategy, either two (level I) or four (level II). The levels divide as follows (LC = Logical Case; CC = Card Case):

(I) CC × Matching; LC × Verification; LC × Falsification.

Table 6.8

The Predictions made on the basis of Matching bias, Verification bias and Falsification in the Selection Task, when (i) Logical case is kept constant, and (ii) Card case is kept constant.

Case	Rule	Matching				Verification				Falsification			
		TA	FA	TC	FC	TA	FA	TC	FC	TA	FA	TC	FC
(i)	*pq*	O	•	O	•	O	•	O	•	O	•	•	O
	p¬q	O	•	•	O	O	•	O	•	O	•	•	O
	¬pq	•	O	O	•	O	•	O	•	O	•	•	O
	¬p¬q	•	O	•	O	O	•	O	•	O	•	•	O
		p	˜p	q	˜q	p	˜p	q	˜q	p	˜p	q	˜q
(ii)	*pq*	O	•	O	•	O	•	O	•	O	•	•	O
	p¬q	O	•	O	•	O	•	•	O	O	•	O	•
	¬pq	O	•	O	•	•	O	O	•	•	O	•	O
	¬p¬q	O	•	O	•	•	O	•	O	•	O	O	•

(II) LC × Matching; CC × Verification; CC × Falsification.

Both levels separate antecedent and consequent predictions, but level I draws no distinction between rule forms, whereas level II separates out affirmative and negative antecedents and consequents. If each rule is treated independently, then a further level (III) of detail can be derived which makes 8 predictions per strategy. At this level whether logical case or card case is kept constant becomes irrelevant. These predictions are readily derivable from table 6.8, however, for perspicuity they are detailed in table 6.9. This table reveals a considerable overlap in predictions between the competing strategies. In the next chapter abstract and thematic versions of the selection task will be reported in which the data was analysed at all the levels introduced here.

Manktelow and Evans (1979) conducted a thematic version of the selection task using the food and drinks material described in 5.4. It was observed in 5.4 that the materials they used were unlikely to facilitate reasoning on the task. First, the task retained the inductive status of the earlier abstract selection task, subjects were asked to determine the truth or falsity of the rule. Second, no action orienting scenario or context was provided to cue the appropriate rule conversion. Third, no rationale was provided for why the subjects should want to perform the action. All the factors which seem to conspire to produce facilitation of falsificatory responding were absent. However, Manktelow and Evans (1979) procedures were true to the existing accounts of the thematic facilitation effect. At the time it was thought that thematic content *per se* was the determining factor and they demonstrated that this simply was not the case. They also replicated the matching bias result using thematic

Table 6.9

Level III predictions (by rule form) for each of Matching, Verification and Falsification in the Selection Task, indicating both Card Case and Logical Case forms .

Rule	Clause	Matching		Verification		Falsification	
		CC	LC	CC	LC	CC	LC
pq	(i) Ant.	p > ˜p	TA > FA	p > ˜p	TA > FA	p > ˜p	TA > FA
	(ii) Consq.	q > ˜q	TC > FC	q > ˜q	TC > FC	˜q > q	FC > TC
p¬q	(iii) Ant.	p > ˜p	TA > FA	p > ˜p	TA > FA	p > ˜p	TA > FA
	(iv) Consq.	q > ˜q	FC > TC	˜q > q	TC > FC	q > ˜q	FC > TC
¬pq	(v) Ant.	p > ˜p	FA > TA	˜p > p	TA > FA	˜p > p	TA > FA
	(vi) Consq.	q > ˜q	TC > FC	q > ˜q	TC > FC	˜q > q	FC > TC
¬p¬q	(vii) Ant.	p > ˜p	FA > TA	˜p > p	TA > FA	˜p > p	TA > FA
	(viii) Consq.	q > ˜q	FC > TC	˜q > q	TC > FC	q > ˜q	FC > TC

material.

Along with Pollard (1981), Reich & Ruth (1982) observed that the materials used in the Manktelow & Evans (1979) selection task experiment were far from realistic, or in their terminology were low in thematic content. In order to check whether more realistic materials would facilitate performance they used rules such as:

(6.7) When it is early Molly serves tea.

None of the matching bias predictions were replicated for this selection task using high thematic content. High thematic content was defined as providing materials which established a coherent non-arbitrary relationship between the concrete terms. For these materials they observed a preponderance of verificatory responses, ie. TA > FA, and TC > FC. They also note that responses on the task are not independent between strategies when considered by rule form (cf. above). There are only six wholly independent responses in the negations paradigm selection task:

(1) Choice of $\neg q$ on pq-rule form = falsifying (F),
(2) Choice of $\neg q$ on $p\neg q$-rule form = verifying (V),
(3) Choice of p on $\neg pq$-rule form = matching (M),
(4) Choice of $\neg q$ on $\neg pq$-rule form = F,
(5) Choice of p on $\neg p\neg q$-rule form = M,
(6) Choice of $\neg q$ on $\neg p\neg q$-rule form = V.

They therefore added the number of responses in each of (1) to (6) to provide a score for each strategy:

Low: 27 (M) > 16 (V) > 13 (F).

High: 19 (V) > 15 (F) > 4 (M).

This appeared to indicate that high thematic content simply placed subjects back where they were in standard abstract versions. Reich & Ruth (1982) interpreted this result to indicate that over and above high thematic content memory cueing was also required to facilitate a falsificatory response. However, Manktelow and Evans procedure failed to include the factors hypothesized in the last chapter to facilitate falsificatory responding. Reich & Ruth's procedure was almost identical. Although they provided a scenario it was not action orienting, ie. the task remained inductive, and no rationale was provided. Given which the standard predictive cycle strategy would be expected, which yields verificatory responses.

Performing the same Reich & Ruth scoring on Evans & Lynch's (1973) data reveals that for the only wholly independent predictions matching and falsification came out equal:

(6.8) 18 (F) = 18 (M) > 9 (V)

Recall that the occurrence of a falsificatory response on pragmatic context theory is a result of attempting to *explain* what is on the other side (cf. taxonomic example, chapter 5). Given a class inclusion relation, ie. a taxonomic constraint, subjects should be able to explain, ie. educt from $\neg q$ to $\neg p$. This suggests that the need to determine a contrast class for negated constituents may facilitate the realisation that given the class inclusion $\neg q$ can be explained by $\neg p$. Moreover, this may serve to partially override the non-taxonomic interpretation suggested by the disjoint expression of the rule. What predictions would this hypothesis make? It would suggest more FC responding for rules with negated constituents, but especially for rules with negative consequents. However, the latter prediction is also made by matching. But there is a prediction not made by matching. First, more falsificatory responses would also be expected for rules with affirmative consequents. Constructing contrast classes would be expected to generally facilitate the class inclusion interpretation. And of course the Reich & Ruth score concentrates just on these cases for falsification, where on average 37.5% of subjects made FC selections on affirmative consequent rules in Evans & Lynch (1973). This contrasts with observations of falsificatory responses on purely affirmative tasks, in Wason's (1969) initial condition only 6.25% of subjects selected $\neg q$ (FC). Similarly, in Beattie & Baron (1988) for pure affirmative tasks (without therapeutic prompts, Wason's initial condition was conducted prior to the therapies) only 8.8% of subjects made $\neg q$ (FC) selections. Bayes' rule indicates that subjects are 4.8 times as likely to turn the FC for affirmative rules when this is in the context of a negations paradigm experiment.

This suggests the possibility that if subjects were provided with materials which facilitated the easy identification of a contrast class, ie. the predicates used formed an antonymic relation and were therefore binary, this may facilitate falsificatory responding in the context of a negations paradigm experiment. However the disjoint expression of the rule licenses a verification strategy in accordance with the predictive cycle. Therefore, although some suppression of matching may be expected, verification would still probably be in evidence. Wason (1969) employed materials in a purely affirmative version of the task which conformed to this suggestion. Each card had a one of two coloured shapes on either side of the cards. Only two colours were used. Take a rule like:

(6.9) If there is not a red square on one side, then there is a blue circle on the other.

"Not red square" could indicate a red circle. So when four cards are presented, ie. blue square, red square, blue circle, red circle, it is ambiguous as to which are the TA and TC and FC instances. A blue or a red circle is as equally not a red square, as a blue square.

However, the subject/predicate structure of natural language means that the negation is normally taken to attach to the predicate. Therefore, shape serves to identify antecedent and consequent instances and colour a binary predicate describing them. Wason (1969) observed no facilitation with these materials in an affirmative version of the task. This provides a secure base line from which to judge the hypothesis that such binary materials should facilitate falsificatory responding in a negations paradigm context. In the next section we will also see a contrast in interpretation for non-taxonomic and taxonomic cases which suggests a difference in consequent card selection task behaviour, hence a thematic version of the selection task was also included in the present experiments.

6.3 A situation theoretic analysis of conditionals containing negations

6.3.1 Taxonomic constraints

The interpretations which are suggested by situation theory for taxonomic and non-taxonomic are different. For a taxonomic constraint, disregarding its unified or disjoint expression, simply defines a class inclusion relation between antecedent and consequent. This is only affected by the presence of negations in the following sense. The presence of a negation serves to identify the relevant contrast class, and then the class inclusions go between these classes once identified. The ease of identifying the relevant contrast class will be determined by pragmatic world knowledge concerning the domains of the predicates. Moreover, as we have observed (cf. chapter 4) some conditionals will express unlikely class inclusions because of pragmatic world knowledge. Relational taxonomic constraints also point to fact that if the antecedent class is vacuous then no decision can be reached concerning the truth or falsity of the constraint. Take the following examples:

(6.10) All black things are ravens.

(6.11) If I don't travel to Manchester, I take the train.

(6.10) clearly violates prior beliefs concerning the domains of the predicates, unless of course the domain is appropriately delimited using a prepositional phrase, eg. all black things *in the gardens at the Tower of London*. Since constraints reflect the structure of the world as individuated, (6.10) would not be considered a good candidate for actuality, and hence may elicit falsificatory behaviour (cf. chapter 4 on truth status effects, Pollard & Evans, 1981). Put another way it violates belief bias. Manktelow & Evans (1979:478) observe that (6.11) is nonsensical. Their observation is not quite that determinate, however. (6.11) can be read as "Wherever I travel other than Manchester, I take the train". This is

perhaps a confusion of negation and falsity. A purported instance of the non-negated rule, in which I didn't travel to Manchester, simply bears neither one way or the other on whether I always travel to Manchester by train (cf. chapter 5, taxonomic example). This is a defective truth table account of the reasoning involved.

For the taxonomic case perming negations simply yields the standard set theoretic interpretations. This is implicit in the situation theoretic formalism for a taxonomic constraint, but for perspicuity the interpretations are simply given set theoretically. Where "P" is the set described by p and "Q" is the set described by q, and "$'$" is a contrast class forming operator and "$\subset$" is *proper subset*:

(6.12) *if p, then q* => $P \subset Q$
(6.13) *if p, then* $\neg q$ => $P \subset Q'$
(6.14) *if* $\neg p$, *then q* => $P' \subset Q$
(6.15) *if* $\neg p$, *then* $\neg q$ => $P' \subset Q'$

Forming a contrast class is determined by pragmatic world knowledge of the taxonomical organisation of predicates, context etc. Even if the domain of objects is fairly well defined initially, other pragmatic factors may further delimit the domain. In the example provided by Reich & Ruth (1982), that Molly didn't serve tea, opens up the domain of drinks as the contrast class, but in a context where she is serving tea, the specific drink *coffee* may be more contextually appropriate. These interpretations are of course constrained by the observation that instances which can not provide an anchor for the antecedent bear neither one way or the other on the rules truth or falsity (cf. chapter 5, and chapter 4: Johnny and the pipes). Hence, these interpretations license the defective truth table view of the truth conditions of the conditional.

6.3.2 Non-taxonomic constraints

The above interpretations apply to taxonomic constraints, where antecedent and consequent are unified by describing either instances of objects or occurrences of single events. When a constraint is non-taxonomic relating disjoint and discrete occurrences of events, then the interpretations above do not apply. In discussing Evans (1983b) above, it was observed that negations may fail to identify a contrast class, sometimes they simply denote the non-occurrence of an event. With regards to taxonomic constraints this is a scope distinction, the negations in a taxonomic constraint attach to the constituents of the relations or parts of the object (cf. chapter 4). But negation can also attach to the whole event, ie. a version of external negation (cf. chapter 2). In taxonomic constraints *discrete* events are described in

antecedent and consequent. Often, a la Evans, the negations attach externally, denying the occurrence of the event. And the relation which holds between these events indicates that the (non) occurrence of the antecedent event is the cause, reason, enablement for the (non) occurrence of the consequent event.

The interpretations for non-taxonomic constraints containing negatives are provided below, "P" stands for the type described by the antecedent *p* and "Q" the type described by the consequent *q*:

(6.16) *if p, then q* -> P => Q
(6.17) *if p, then ¬q* -> P =| Q
(6.18) *if ¬p, then q* -> P =| Q, Q =| P
(6.19) *if ¬p, then ¬q* -> P => Q, Q => P

(6.16) and (6.17) require no further motivation. In (6.18) and (6.19), the constraints are not logically conjoined because these are operative constraints. This is like specifying the meaning of the conditional by the inference rules which determine it logical behaviour. Similarly (6.18) and (6.19) say that these relations determine those conditionals information gaining behaviour. For (6.18), take the following example:

(6.20) If the boss doesn't want to see me, I'll be home in time for dinner.

This could be paraphrased as the only thing which ever prevents me being home in time for dinner is if the boss wants to see me, ie. P =| Q. This reflects the fact that pragmatically the boss wanting to see me is an impediment to my arriving home for dinner. The antecedent negation appears to be marking exclusivity, ie. not only would the boss wanting to see me prevent me being home, it is the only thing which would do this. This carries the information that if I arrive home in time for dinner, then the boss didn't want to see, Q =| P. A similar argument applies to (6.21):

(6.21) If I don't finish my work, I won't be home in time for dinner.

However, in (6.21) finishing my work is an enablement for being home in time for dinner, ie. P => Q, and the exclusivity again indicates that whenever I arrive home for dinner, then I finished my work, Q => P.

The truth conditions licensed by (6.18) are equivalent to exclusive-OR (XOR), ie. it is true just in case either the boss wants to see me and I make it home in time for dinner (FT) or the boss doesn't want to see me and I don't make it home in time for dinner (TF), it is false otherwise. However, (6.19) is just the opposite, it is true just in case either I don't finish my work and I don't make it home (TT) or I finish my work and make it home (FF),

it is false otherwise. (6.16) and (6.17) license the same truth conditions as (6.12) and (6.13) respectively.

However, when looking at the eductions licensed by each of (6.18) and (6.19), each constraint still possesses a background type, other contextual factors have to be fixed. For example, although the only condition relevant to my promise in (6.21) is whether or not I finish my work, if I do finish my work, but the bus breaks down and I don't make it home, then although the claim is false I didn't lie. This is not pertinent to the truth conditions of the claim but are highly salient relative to the eductions which can be performed. All other things being right if I finish my work an eduction to my being home in time for dinner is sound. Also I may not finish my work but be home in time for dinner, because one of my daughters has had an accident, then although the claim is false, again I didn't lie. All other things being right my wife can educt to my having finished my work from my being home. This is just the situation described in the non-taxonomic example in chapter 5. (6.18) and (6.19) explicitly state that a good explanation of why I'm home is either the boss didn't want to see me or I finished my work respectively. But simply because this is explicit, rather than implicit, as in my hall lights example, does not mean any further information gaining eductions are licensed. This could account for why so few subjects ever turn all the cards, ie. treat the conditional as a material biconditional.

The contrasting interpretations for taxonomic and non-taxonomic constraints have been motivated by appeal to intuitions. However, they cna be derived directly from the differing anchoring conditions and contextual assumptions which attach to these constraint types. Initially a uniform interpretation can be provided for conditionals containing negations, where "¬" = dual:

(i) P => Q
(ii) P =| Q
(iii) ¬P => Q
(iv) ¬P =| Q

Given these uniform interpretations, the contrasts result from considering the different anchoring conditions. Recall that taxonomic constraints induce a restriction on the anchor for the indicating type in the indicated type. Taking a rule, for example:

(6.22) If triangle right, then square left

and allowing the negations to perm through, yields the following interpretations:

P: <<*Card_pair*, *right:* x, *left:* y; *1*>> & <<*Triangle*, x; *1*>>

Q: *<<Square, y; 1>>*

If triangle right, not square left:

P: *<<Card_pair, right:* x, *left:* y; *1>>* & *<<Triangle,* x; *1>>*

Q: *<<Square,* y; *0>>*

If not triangle right, square left:

P: *<<Card_pair, right:* x, *left:* y; *1>>* & *<<Triangle,* x; *0>>*

Q: *<<Square,* y; *1>>*

If not triangle right, not square left:

P: *<<Card_pair, right:* x, *left:* y; *1>>* & *<<Triangle,* x; *0>>*

Q: *<<Square,* y; *0>>*

Relative to the anchoring conditions for taxonomic constraints, each states that any anchor for the indicating type P must be restricted such that it is also anchors y to objects which are (not) squares. The contrast classes for each dual will be contextually determined, eg. in construction tasks it will be given by the card array.

For non-taxonomic constraints similar contextual effects will also be operative. It was mentioned in chapter 2 that the interpretation of a condionla sentence as asserting the existence of a constraint will also depend on the other constraints which may be operative in a situation. Non-taxonomic constraints do not simply induce a restriction on the anchors for the indicating type in the indicated type. The acnhoring conditions indicate that discrete events are related by a higher order relation. The examples provided by (6.20) and (6.21) will be used to illustrate the appropriate interpretations.

P: *<<Is_being_worked_on_by,* l, x, y; *1>>* & *<<Me,* y; *1>>* & *<<Finishes,* y, x; *1>>*

P′: *<<Wants_to_see,* l, y, x; *0>>* & *<<Me,* y; *1>>* & *<<Boss,* x; *1>>*

Q: *<<Home,* l′, y; *1>>* & *<<Dinner_time,* l″; *1>>* & *<<Overlaps,* l′, l″; *1>>* & *<<Follows,* l′, l; *1>>*

All negated versions will not be presented. As with all negative soas the focus of the negative is ambiguous unless disambiguated via intonation. The first conjunct of the indicating types, P and P′, and the indicated type, Q, have additional restrictions placed on them by the subsequent conjuncts. A negation attaches to the first conjunct but it can focus on the

restrictions. So what is the focus of "I don't finish my work"?. Well it could be someone else finishes it, ie. the negation attaches to the second conjunct. In which case (6.21) would be interpreted as, "If someone else finishes my work, I won't be home in time for dinner". This is pragmatically anomalous, since now I don't have to do the work I am free to be home for dinner. The consequent pragmatically affects the interpretation of the antecedent. The pragmatically felicitous interpretation would appear to be where the negation attaches to the third conjunct, ie. the work is not finished by me. In the indicated type, a negation attaching to the fourth restriction would be anomalous given knowledge of the temporal sequencing of events. Attaching to the third, yields the contrast class of a time later than dinner time. So focus can be felicitous or not dependent on pragmatic world knowledge. However, wherever the negations felicitously focus the interpretations licensed are as indicated in (6.16) - (6.19).

(6.16) has already been extensively discussed. In (6.17), this is the only eduction licensed because there may be events other than P′ which could preclude my arrival home in time for dinner, eg. the departmental meeting went on late. In (6.18) and (6.19), the appropriate contextual assumptions need to be specified. This is required to identify the force of the antecedent negation. The antecedent negation in (iii) indicates that within a situation everything but P involves Q, in which case given Q the only thing that would not be expected is P, ie. Q =| P. But equally if everything but P involves Q, and Q =| P, then P =| Q. So, if I assert that anything else that happens apart from the boss wanting to see me (within my control) will involve my being home, then if I arrive home an eduction to the boss not wanting to see me is sound. Equally, if anything other than the boss wanting to see me will involve my being home in time for dinner, then the boss wanting to see me is the only thing (within my control) which could preclude my being home in time for dinner. A directly analogous argument applies to (6.19). If my not finishing my work (cf. above on focus) will preclude my being home in time for dinner, then my being home in time for dinner must mean I finished my work. Equally if I finish my work, given that not finshing it is the only thing which will preclude my being home, then I will be home in time for dinner.

These contrasting interpretations for conditionals containing negations describing taxonomic and non-taxonomic constraints will function to motivate the experimental hypotheses in the next section.

6.4 Introduction to the experiments

6.4.1 Construction tasks

The interpretations provided in the last section do indeed predict an interaction between negation and the conditional, *but* only for non-taxonomic constraints. However, it has already been observed that for the standard abstract selection task the disjoint expression of the rule forces a non-taxonomic interpretation (cf. chapters 4 & 5, and Wason & Green, 1984). The rule used in Evans (1972) had a disjoint expression and in Evans (1983b), subjects were told that the numbers and letters occurred on each side of the card. On this basis it could be hypothesized that the occurrence of matching in subjects initial response data in Evans (1972) was a result of adopting the XOR interpretation for the *if* $\neg p$, q rule form which is appropriate to the non-taxonomic case. This is borne out in the total responses where it appeared that subjects were interpreting this rule as XOR. As indicated above, however, this interpretation is only appropriate given the kind of content in (6.20) and (6.21) where discrete events are being related. This could permit a critical test of whether this result is due to an albeit contextually apposite, misinterpretation of this rule form. If subjects conducted a thematic version of the task using materials appropriate to the non-taxonomic interpretation, then if the FT instance was also subjects initial falsifying construction, this would implicate the misinterpretation hypothesis. The occurrence of FT as subjects initial falsifying construction could be considered the result of the processes involved in constructing that interpretation. Alternatively, if in the thematic version subjects initial construction was TF as falsifying, subjects abstract construction task behaviour would have to be put down to other factors, ie. either matching or processing negations. Moreover, if the differences were due to interpretational factors then a similar effect would have been expected for the *if* $\neg p$, $\neg q$ rule form. However, although there was a tendency to construct FF as falsifying, this did not predominate over TF in the initial responses. This response is, however, inconsistent with an interpretational shift to material equivalence for this rule. *But* it is consistent with matching, FF is the double match case for this rule, it nonetheless did *not* predominate.

Conducting two such construction tasks will also serve to test the hypothesis concerning subjects rule interpretations. For the thematic material where a non-taxonomic constraint is the appropriate interpretation, responses which accord with XOR and material equivalence would be expected for negative antecedent rules. And for these rules more of these interpretations would be expected in the thematic case over the abstract case. Moreover, if explicit negations were included in the thematic task, the appropriate thematic content

would be expected to override the general shift towards a material equivalence interpretation observed in Evans (1983b). That is, more verifying constructions would be expected for FF instances and falsifying constructions on FT instances for negative antecedent rules over affirmative antecedent rules. Moreover, a perceptual matching strategy would predict a general shift for all rule forms. If the differences just adduced are found, then perceptual matching can be rejected as the general strategy being adopted for the thematic material containing negations.

Testing for the occurrence of matching is relatively easy. Matching results on these tasks are reported by group data pooled over the rules. In general this is a sound procedure precisely because it makes symmetrical predictions across the response record. However, this is also a weakness of the matching bias hypothesis, especially one which attributes the phenomenon to an early processing heuristic. The interpretational account makes asymmetric predictions relative to rule form and instance. If most rules fail to display matching predictions, then it is generally unsound to attribute the exception to matching, rather than say a difficulty in processing antecedent negations, for example. The only asymmetry predicted by matching is between 0,1 and double mismatching cases. Again if symmetrical responding is observed for most cases, or it goes in the other direction, then again inferring that matching is responsible for the exception is unsound. Hence, if the only significant matching result is due to the initial construction of the FT falsifier for the *if* $\neg p$, q rule form, then it can be reasonably concluded that a general matching strategy is not implicated.

In Evans (1972) it was argued that the salience of TT and TF cases probably override the effects of matching. However, this argument is no longer available to Evans given the serial processes he suggests are implicated in subjects responses. Evans (1972) position was more reasonable than the two stage model of Evans (1983b). In Evans (1972), matching was simply hypothesized to facilitate or inhibit the construction of the logically correct TF falsifier. In the two stage model the processes responsible for the phenomena are actually hypothesized to block the interpretational process.

As computed by Evans (1983b), the matching indices for each task would be expected to alter in the direction he predicts simply in virtue of the predicted interpretational shift for negative antecedent rules. In the results indices will therefore be computed correcting for predicted interpretational changes. The indices Evans computes are noteworthy for the fact that the double mismatch case enters into each computation, in this regard they are not *independent* measures of matching. They are also inappropriately named, since they actually

measure the occurrence of irrelevants, be they due to matching or otherwise. A "pure" matching index will be computed which corrects for interpretational changes, in a manner similar to Evans logic index. Once corrected it would be predicted that there would be more pure matching for negative antecedent rules in the abstract task, since it would appear that it is antecedent negation which is responsible for the anomalous results.

Evans had different groups of subjects perform the two versions of the task in Evans (1983b). In order to function as a more effective test of the interpretational changes hypothesized it was felt appropriate to have subjects act as their own controls, therefore all subjects conducted two versions of the construction task, one abstract and one thematic. Transfer effects in these tasks are rarely observed (cf. Beattie & Baron, 1988). However, they were tested for in the results. The materials used are introduced in the experimental reports in the next section. The thematic materials used were those in (6.20) and (6.21), the negations were not wholly explicit but in their morphologically abbreviated form as were the statements on the cards in the array.

6.4.2 Selection task

It was suggested above that constructing contrast classes in the context of a negations paradigm experiment may alter subjects abstract selection task performance towards more falsificatory responding when purely binary materials are used. To test this hypothesis the same subjects as those above were given an abstract selection task to perform. The analysis of non-taxonomic constraints predicts that although subjects may interpret the rules as equivalence in the construction tasks they will nonetheless verify in an inductive version of the selection task. Again to test this hypothesis, having subjects act as their own controls and perform both tasks was deemed appropriate. This meant that each subject performed 4 tasks, an abstract and thematic construction task and an abstract and thematic selection task. The orders were systematically permuted such that no abstract or thematic task followed in succession. This was to avoid any possible transfer effects.

The exhaustive levels of analysis for a selection task will also be carried out in order to obtain as determinate an answer as possible concerning the strategy which best summarises subjects response profiles. As observed above different measures seem to imply different or conflicting summary strategies. Evans & Lynch (1973) initially appeared as conclusive evidence for matching. However, the Reich & Ruth scores, although wasteful of data, indicated ambiguity between matching and falsification. With this in mind, *Pollard indices* will

also be computed for each of matching, verification and falsification. I follow Manktelow & Over (personal communication) in naming this index after Pollard, however, it is simply an extension to selection task data of the matching indices introduced by Evans (1983b).

Summary

This chapter has critically reviewed the data from Evans' negations paradigm in the selection task. Several experimental hypotheses have been derived which suggested the experiments to be reported and discussed in the next chapter.

Chapter 7: The Experiments

7.1 Introduction and overall design

This chapter reports the experiments introduced in chapter 6. Every subject conducted each experiment. The experimental order was systematically varied such that no two tasks, selection or construction, were juxtaposed. This was designed to prevent the possibility of transfer between related task. This yielded only 2 possible task orders. Combining with all possible permutations of task type, abstract or thematic, yielded 8 possible task orders. There were 24 subjects, so three subjects received the same task orders which were assigned randomly. An analysis for transfer effects between the various tasks was conducted and will be reported in the results section of experiment 1. Anticipating those results, no transfer effects were observed. Taken together with the richness of the data to be reported from each task, treating each as a separate experiment rather than different conditions of the same experiment was deemed appropriate.

7.2 Experiment 1: Abstract Construction Task

Introduction

This experiment was a replication of Evans (1972) with only minor procedural differences. The complete introduction is given in Chapt 6.

Method

Subjects

24 undergraduate psychology students from the University of Edinburgh served as subjects on an unpaid volunteer basis. All subjects were tested individually on all conditions.

Design

Subjects were required to construct both verifying and falsifying instances of each of four conditional rules of the form: *if p, then q; if p, then* $\neg q$*; if* $\neg p$*, then q; if* $\neg p$*, then* $\neg q$, as in Evans (1972). Each subject was given one of the 4! permutations of presentation order of the rules. Evan's (1972) divided subjects into two halves, one half constructed verifying cases before falsifying cases on each rule, the other half constructed falsifying cases before verifying cases. It was felt that this method of counterbalancing could induce a "strategy set". The more regularity in the task the more subjects may be inclined to adopt a single strategy and ignore the changes in semantic/pragmatic content. So, for each subject the verify/falsify order was systematically varied such that for each rule, the same order could occur in succession only once. This still meant that half of all subjects responses were in each order, so counterbalancing the group data. It also meant that each subject's data was individually counterbalanced within the task. Rule presentation orders were randomly assigned.

Task Materials

The task materials were the same as in Evans (1972), apart from one minor detail. Evans's used a 4 × 4 array of stimulus cards which depicted various coloured shapes. The shapes were: circle, triangle, cross, and square; the colours were: red, yellow, green and blue. In the present experiment a diamond was used in place of the cross. The general form of the rules employed is illustrated below:

(7.1) If there is a red diamond on the left, then there is not a green circle on the right.

The lexical material was varied randomly between between rule forms. Subjects task was to select a pair of cards from the array which either verified or falsified the rule. For example, a subject may place a red diamond on the left and a blue circle on the right to verify. This is equivalent to constructing the TT logical case (*if p,* $\neg q$ card case) as making the rule true.

Procedure

Each subject was presented with the 4 × 4 array of single sided colours and shapes stimulus cards. The following typed instructions were then given to the subject:

(7.2) "You will be presented with a series of rules which always assume that two of the figures before you have to be placed side by side. Your task will be to select two of the figures from the array and place them in such a way as to make a given rule true, or in such a way as to make a given rule false, according to the instruction. If you have any questions please ask them now and not after you have started on the problems."

When a subject had finished reading the instructions, he was told that they could keep the instructions beside them to refer to. The four rules were then presented one at time on a typed sheet in the order of presentation assigned, and from the set of materials assigned. Dependent on the verify/falsify order assigned on a rule, a subject was asked to make a selection which made it true or made it false. For each verify or falsify instruction, once a subject had provided one solution, they were asked whether there was any other selection which made the rule true (false) This was continued until the subject said there were no more. Generalisations such as "a red diamond on the left with anything other than a green circle on the right would make the rule true" were permitted. Thus, an exhaustive series of verifying and falsifying constructions was obtained for each rule.

Results

(i) Transfer effects

The Cochran Q test was used. The dependent variable employed was correct or incorrect. This was the same procedure carried out by Beattie & Baron (1988) in a similar multi-task design. For the selection tasks "correct" was defined as a TA and TC selection, in accordance with pragmatic context theory, and for the thematic construction tasks "correct" = TT to verify and TF to falsify on affirmative antecedent rules, and TT & FF to verify and TF and FT to falsify on negative antecedent rules. For the abstract construction task "correct" = TT to verify and TF to falsify. There were four tasks and four rules leading to 16 serial positions. Cochran's $Q(15) = 16.612$, which is not significant.

It was possible that this result was a function of pooling the data across tasks. Therefore, the data was broken down by tasks and similar tests performed. Cochran's Q, nor any other

test statistic, is wholly appropriate to the data when analysed by tasks since in each task position the same six subjects conducted the task on all four rules. Hence, the design, when taking task/rule position as the dependent variable, is neither fully repeated measures nor fully independent. However, it could be argued that the subjects are a sufficiently matched sample for Cochran's Q to be appropriate; all were undergraduate psychology students at the University of Edinburgh. The results of the tests by task were, Abstract selection: $Q(15) = 11.61$, n.s.; Thematic selection: $Q(15) = 22.35$, n.s.; Abstract construction: $Q(15) = 13.99$, n.s.; Thematic construction: $Q(15) = 24.88$, n.s. Since, there were no significant changes in the distribution of correct responses either in the pooled data or by task it can be concluded that there was no tendency for subjects to perform better on the tasks as a function of the order in which they were performed. Therefore, each experiment is reported individually.

(ii) Initial and Total Response Analysis

Table 7.1 shows the frequency with which subjects gave each truth table case on their initial verification and initial falsification of each rule. Subject's initial responses were an almost direct replication of Evans (1972). The TT construction was almost universally chosen as the only verifying case. This was consistent with Wason's (1966) "defective" truth table account. The initial choice of the FT construction as a falsifier for the *if* $\neg p$, q rule form was also replicated more strongly than Evans (1972). It was this initial choice which was observed to be responsible for the confirmation of Evans matching bias predictions. The prediction that there would be fewer TF falsifiers constructed for rules with negative antecedents than affirmative antecedents was significantly replicated, using one-way sign tests ($p < 0.0005$). The prediction that there would be more TF falsifiers

Table 7.1

The frequency with which subjects gave each truth table case on their initial verification and initial falsification of each rule in the Abstract Construction Task.

Rule	Verification				Falsification			
	TT	TF	FT	FF	TT	TF	FT	FF
(i) If p, then q	23	1	0	0	1	18	3	2
(ii) If p, then ¬q	23	0	0	1	0	23	0	1
(iii) If ¬p, then q	23	1	0	0	0	3	19	2
(iv) If ¬p, then ¬q	23	0	0	1	0	9	2	13

constructed for rules with negative consequents was also significantly replicated ($p < 0.005$).

It was argued in Chapter 6 that testing for these predictions only in subject's initial response data was the artefactual cause of their significance. Table 7.2 shows the total frequency with which each subject gave each truth table case on verifying and falsifying each rule. Subject's total response data revealed the following results. The prediction of fewer TF falsifiers for rules with negative antecedents was still significant ($p < 0.001$. 1-tailed sign test). This prediction was *not* significant in the total response data in Evans (1972). The reason for this can be seen from Table 2. There has been a shift in predominance as the main falsifying response from TF to FF for the *if* $\neg p$, $\neg q$ rule form between Evans (1972) and the current abstract version of the task. This result may be due to the presence of contrastive information (cf. Discussion). The prediction of more TF falsifiers for rules with negative consequents was not significant in the total response data, replicating Evans (1972). In general, for subject's total responses, there was a failure to replicate parts of Evans (1972) (cf. Discussion).

(iii) Psychological truth tables

Table 7.3 shows the result of plotting the frequency of true, false and "irrelevant" (non-constructed) classifications of each truth table case. There were important differences between the present abstract version and Evans (1972) which all occurred for rules with negative antecedents. For the *if* $\neg p$, q rule form, the FF instance shifted radically from being treated as true in Evans (1972) to being treated as irrelevant. For the *if* $\neg p$, $\neg q$ rule form, the TF case moved from false to being treated as irrelevant.

Table 7.2

The total frequency with which each subject gave each truth table case on verifying and falsifying each rule for the Abstract Construction Task

Rule	Verification				Falsification			
	TT	TF	FT	FF	TT	TF	FT	FF
If p, then q	23	1	0	2	1	21	6	4
If p, then ¬q	24	0	0	3	0	23	4	1
if ¬p, then q	23	1	0	2	0	16	21	2
if ¬p, then ¬q	24	0	0	1	0	11	5	14

Table 7.3

The frequency of true, false and "irrelevant" (non-constructed) classifications of each truth table case (N = 24; modal response is in italics).

Rule	Truth Value	Truth Table Case			
		TT	TF	FT	FF
(i) If p, then q	T	*23*	1	0	2
	F	1	*21*	6	4
	?	0	2	*18*	*18*
(ii) If p, then ¬q	T	*24*	0	0	3
	F	0	*23*	4	1
	?	0	1	*20*	*20*
(iii) If ¬p, then q	T	*23*	1	0	2
	F	0	*16*	*21*	2
	?	1	7	3	*20*
(iv) If ¬p, then ¬q	T	*24*	0	0	1
	F	0	11	*5*	*14*
	?	0	*13*	*19*	9

Consistent with Evans (1972), the TT case was significantly more often constructed as true than false for all rules ($p < 0.0005$, one-tailed Binomial test). The TF case was significantly more often constructed as false than true for all rules ($p < 0.0005$, one tailed Binomial test). In reporting these results Evans (1972) tied the analysis to the matching bias hypothesis. Matching was hypothesized to suppress the construction of certain cases, "which involved mismatching or altering named values, thus increasing their apparent irrelevance" (Evans 1972:197). The occurrence of the inferred "irrelevant" value was thereby firmly tied to the matching bias hypothesis. Hence, in testing for the significance of each response category, ie. rule form by truth table case, Evans (1972) did not test for differences between the irrelevant value and the other two truth values but only for the direction of classification, true or false. Since the source of irrelevants is the subject of test in these experiments, differences between true, false *and* irrelevant response categories were computed for each rule form × truth table case. However, this only applies to the false antecedent instances (FT & FF) because the "irrelevant" category was significantly more subscribed to than the other truth values only for these instances. (For TF in the *if* $\neg p$, $\neg q$ rule form, "irrelevant" predominated, but not significantly, over false ($p = 0.838$, 2-tailed Binomial test). For the *if* $\neg p$, q rule form, TF was constructed significantly more often as false than irrelevant ($p = 0.047$, 1-tailed Binomial test). All tests are 1-tailed Binomial tests where the direction is predicted by the defective truth table/taxonomic constraint interpretation and two-tailed

predicted by the defective truth table/taxonomic constraint interpretation and two-tailed otherwise.

For the *if p, q* rule form there were significantly more FT cases constructed as false than true ($p = 0.016$, 2-tailed), but FT was also significantly more often treated as irrelevant than false ($p = 0.011$, 1-tailed). FF was more often constructed as false than true, but not significantly ($p = 0.688$, 2-tailed), whereas it was significantly more often treated as irrelevant than false ($p = 0.002$, 1-tailed). For the *if p, ¬q* rule form, although FT was more often constructed as false than true, this was not significant ($p = 0.124$, 2-tailed) whereas it was treated as irrelevant significantly more often than false ($p < 0.001$, 1-tailed). FF was more often constructed as true than false, but not significantly ($p = 0.624$, 2-tailed) whereas it was treated as irrelevant significantly more often than true ($p < 0.0005$, 1-tailed). For the *if ¬p, q* rule form, FT was significantly more often constructed as false than either true ($p < 0.001$, 2-tailed) or irrelevant ($p < 0.001$, 2-tailed). FF was constructed as true and false in equal numbers, but was treated as irrelevant more often than both true ($p < 0.0005$, 1-tailed) or false ($p < 0.0005$, 1-tailed). For the *if ¬p, ¬q* rule form, FT was constructed as false more often than true but not significantly ($p = 0.062$, 2-tailed), whereas it was treated as irrelevant significantly more often than false ($p = 0.003$, 1-tailed). FF was constructed as false significantly more often than true ($p < 0.001$, 2-tailed) and more often than irrelevant, but not significantly ($p = 0.404$, 2-tailed). These results are consistent with pragmatic context theory except for the TF instance in the *if ¬p, q* rule form, which replicates Evans (1972) and the FF instance for the *if ¬p, ¬q* rule form which does not.

From table 7.3, the three valued truth tables in table 7.4 were extracted. 14 out of 16 (87.5%) of these truth table entries reflect significant differences from both remaining possible response categories.

Table 7.4

Psychological truth tables extracted for each rule form.

	Rule-Form							
	if p, q		if p, ¬q		if ¬p, q		if ¬p, ¬q	
	q		q		q		q	
	0	1	0	1	0	1	0	1
p 0	?	?	?	?	0	1	1	?
p 1	0	1	1	0	?	0	?	0

(iv) Defective truth tables?

Evans (1982) re-analysed the data from Evans (1972) adding further weight to a response bias interpretation of subjects irrelevant responses (cf. Chapt 6). It was observed that this analysis relies on the assumption that negations do not interact with the conditional to affect subjects interpretations. However, it has been argued that negation does indeed interact with the conditional, *but* indirectly via the relations implicitly asserted to exist between antecedent and consequent. For the simple class inclusion relation pragmatic context theory would predict responses which accord with the defective truth table. The results of the same analyses as Evans (1982) are reported below.

Table 7.5 pools the results over the four rules classified by (a) logical case and (b) matching case. From table 7.5 it can be seen that in the present data the defective truth table account received even stronger corroboration than in Evans (1972). The analysis by logical case clearly indicates that irrelevant was the modal response for cases with false antecedents. For matching case, a *trend* for more irrelevants was not observed. Although mismatching cases did in general yield more irrelevants there was no increase for double mismatches over single mismatches where it would be most expected on the basis of the matching bias hypothesis. This is even more clearly demonstrated by looking to the frequency of irrelevants by mismatching case within rules. Table 7.6 shows the frequency of irrelevants by rule as a function of the number of mismatches. None of the expected

Table 7.5

Results pooled over the four rules and classified by (a) Logical Case and (b) by Matching Case. Results are percentage frequencies (N = 24).

Case	Classification		
	True	False	Irrelevant
(a) Logical			
TT	*98*	1	1
TF	2	*74*	24
FT	0	*37.5*	*62.5*
FF	8	22	*70*
(b) Matching			
p, q	25	61	14
$p, \neg q$	28	29	43
$\neg p, q$	27	19	54
$\neg p, \neg q$	28	25	47

Table 7.6

The frequency of "irrelevants" (non-constructed) classifications for each rule and mismatching case.

Rule	Number of Mismatches			
	0	1	1	2
(i) If p, then q	0	2	18	18
(ii) If p, then ¬q	1	0	20	20
(iii) If ¬p, then q	3	1	20	7
(iv) If ¬p, then ¬q	9	13	19	0
	13	16	77	45

increases in irrelevants predicted by matching occur within rules. The failure to observe any increase, but rather decreases, between 1 and 2 mismatches severely questions the viability of the matching hypothesis. Matching would predict a step function, with sharp rises between 0 and 1 mismatching cases and between 1 and 2 mismatching cases, with a plateau for the two single mismatches. *None* of the profiles by rule type reveal this pattern.

Table 7.7 shows the result of plotting the frequency of true, false and irrelevant responses for rule forms/logical cases which constitute a double match, as in Evans (1982). Table 7.7 is a direct replication of Evans (1982/1972), apart from the more clear cut falsifying classification of the FF instance for the *if* ¬*p*, ¬*q* rule. However, there is no reason to suppose that this is not due to a another trend, ie. for more *erroneous* irrelevant responding as a function of the number of negations in the rule, when matching is held constant. This is consistent with the failure to observe the predicted trend for irrelevants as a function of matching and the complementary observation of the predicted response profiles by logical case. However, it is only appropriate if it can be established that subjects are

Table 7.7

The percentage frequency of construction of each logical case on rules where they constitute a double match, p, q. (N = 24).

Logical Case	Rule	% Frequency True	False	Irrelevant
TT	If p, then q	96	4	0
TF	If p, then ¬q	0	96	4
FT	If ¬p, then q	0	87.5	12.5
FF	If ¬p, then ¬q	4	58	38

misinterpreting the rule as a non-taxonomic constraint due to the disjoint expression of the rule. To determine this will have to await the discussion of the thematic construction task.

(v) Evans' matching indices

Antecedent and consequent matching indices were calculated for each subject (Evans, 1983b). AMI had a mean of 1.583 and a standard deviation of 1.412, CMI had a mean of 0.917 and a standard deviation of 1.288. If an index is positive then this indicates the presence of matching, by the criterion laid down by Evans (1983b). A significant majority of subjects produced positive indices on both AMI (19 +, 3 -, $p < 0.0005$, 1-tailed Binomial test) and CMI (18 +, 4 -, $p = 0.015$). Given the failure to observe, by inspection, any of the predicted increases in irrelevants between mismatching cases, the significance of these results is suspect. The only reason for their significance is that they conflate two sources of irrelevant responding, partial interpretation and matching. Below an index is derived to unconflate these sources of irrelevants.

Evans (1983b) provides the means and standard deviations of the matching scores he obtained which permits a comparison. However, each of his subjects performed the task twice, so to obtain comparable statistics the mean and the standard deviation were doubled. AMI was lower than observed in Evans (1983b) I-group, but not significantly. However, CMI was significantly lower in the present experiment (unequal Ns, independent samples *t*-test, $t = 2.36$, 62 df., $p < 0.05$, 2-tailed).

The results seem to support Evans contention that their is a significant matching bias observed in these tasks. However, pragmatic context theory represents a competence model of interpretation which *predicts* the occurrence of irrelevants. This is a different contention than the claim that the observation of irrelevants can be explained by a defective truth table in the logical component. This may or may not be the case. However, since the competence model provided by pragmatic context theory predicts the occurrence of irrelevants for false antecedent instances *and* Evans (1983b) concedes that irrelevants occur for this reason (although he did not correct for it) an appropriate correction needs to be applied.

The simplest correction which begs the fewest questions against the matching hypothesis as possible is to allow that where pragmatic context theory fails to predict irrelevants, then these can be attributed to matching. Where irrelevants are predicted, ie. for false antecedent instances, the following strategy should be adopted. For each rule form, there are two

irrelevant predictions, one of which matches once more often than the other, eg. for affirmative antecedent rules one is a 2-mismatch case the other a 1-mismatch case, for negative antecedent rules, one is a 1-mismatch case, the other a 0-mismatch case. If the number of irrelevants is tied between mismatching cases then matching can be assumed not to be present, since the extra mismatch has had no effect on subjects' behaviour. Alternatively if there is an irrelevant only for the extra mismatch then this should be assigned to matching. For example, take subject 21's response record for the *if* $\neg p$, $\neg q$ rule form by matching case (a "1" = irrelevant): 0 (pq), 1 ($p\neg q$), 1 ($\neg pq$), 0 ($\neg p\neg q$), both (pq) & ($p\neg q$) are predicted to be irrelevant, but only ($p\neg q$) is found to be assigned irrelevant, since this is in the direction predicted by matching, subject 21 is assigned a *pure matching* score of 1. Moreover, since an irrelevant also occurred ($\neg pq$) where it was not predicted, he gets another matching score of 1, totalling 2. Matching scores can then be calculated for each subject for each rule. Two dichotomous categories are created, non-matchers (0 matching score) and pure matchers (+ve matching score), hence the Binomial test was employed to assess matching levels within rule forms. This facilitated a test of the hypothesis that there would be little matching for affirmative antecedent rules, if subjects are either (i) having problems with processing antecedent negation or (ii) a non-taxonomic interpretation had been adopted due to the disjoint expression of the rule. Matching would not hypothesize any differences between rule forms.

The results by rule form were as follows. For the *if* p, q rule form a significant majority of subjects were non-matchers (17 0, 7 +, $p = 0.032$), the pure matching score (PMS) had a mean (M) of 0.292 and a standard deviation (SD) of 0.455. This result was significantly replicated for the *if* p, $\neg q$ rule form (21 0, 3 +, $p < 0.0005$; PMS: M = 0.125, SD = 0.331). For the *if* $\neg p$, q rule form the majority of subjects were pure matchers (3 0, 21 +, $p < 0.0005$; PMS: M = 1.125, SD = 0.612). However, for the *if* $\neg p$, $\neg q$, although there were more pure matchers than non-matchers, the result of the binomial test was not significant (9 0, 15 +, $p = 0.154$; PMS: M = 1.042, SD = 0.889).

There was no significant difference in the occurrence of pure matching between the *if* p, q and *if* p, $\neg q$ rule forms ($t = 1.699$, n.s). However, there was a highly significant increases in pure matching over the *if* p, q rule form for the negative antecedent rule forms: *if* $\neg p$, q ($t = 4.275$, $p < 0.0005$) and *if* $\neg p$, $\neg q$ ($t = 4.097$, $p < 0.0005$). There were also highly significant increases in the occurrence of pure matching over the *if* p, $\neg q$ rule form for negative antecedent rule forms: *if* $\neg p$, q ($t = 5.874$, $p < 0.0005$) and *if* $\neg p$, $\neg q$ ($t = 5.102$, $p < 0.0005$). There was no significant difference in the occurrence of pure matching between the negative antecedent rules ($t = 0.358$, n.s.).

As expected from pragmatic context theory, the majority of subjects on affirmative antecedent rules are non-matchers. This is reversed for the negative antecedent rules where the majority display some pure matching behaviour. Moreover, the differences between affirmative and negative antecedent PMSs are nearly all significant. However, within affirmative and negative antecedent rules pure matching levels are not significantly different. This argues strongly that the anomalous results and hence the occurrence of pure matching for negative antecedent rules is not due to any generalisable matching strategy. Rather, either (i) a contextually apposite disjoint, non-taxonomic interpretation has been adopted, or (ii) the anomaly is due to difficulties in processing antecedent negations.

Discussion

These results argue strongly against attributing the observation of anomalous responses on this task for negative antecedent rules to a generalised matching strategy. Matching makes essentially symmetrical predictions which are not borne out in the data. The only significant pure matching result was obtained for the *if* $\neg p$, q rule form. However, there was no significant differences between this rule form and the *if* $\neg p$, $\neg q$ rule form. This appears to suggest that it is processing of antecedent negations which is responsible for the anomalous result, since the *if* $\neg p$, $\neg q$ results do *not* conform to the non-taxonomic interpretations for this rule form. However, although it would appear that matching is definitely not responsible, at least in the form suggested by Evans (but cf. below), the negations/conditional interaction hypothesis is not totally ruled out. This can only be determined by the results from subjects initial interpretations in the thematic version of the task.

There were two interpretational shifts between the Evans (1972) data and the present results. This represents a small change in response profiles which could easily be put down to population differences. However, there is a another factor present in this experiment which was absent in Evans (1972) which could account for the variation. The class inclusion relationship use in the present experiment invites different inferences and engages different modes of thought from the enablement relation in the thematic version (cf. Chapt 6). This is the hypothesis these experiments are investigating. If the encoded message in the conditionals is unclear, then varying the encoded information between the abstract and thematic variants, while retaining task structure, could allow subjects to use *contrastive* information. This was possible in the present experiment because subjects performed both an abstract and a thematic version of the task. Hence, a possible explanation for the discrepancy between Evans (1972) and the present experiment was that in the original

study this contrastive information was not available.

7.3 Experiment 2: Thematic Construction Task

Introduction

This experiment was a replication of Evans (1972) but using thematic materials. The complete introduction is given in Chapt 6.

Method

Subjects

The same 24 undergraduate psychology students from the University of Edinburgh served as subjects on this experiment as in Experiment 1. All subjects were tested individually on all conditions.

Design

The design of this experiment was the same as Experiment 1.

Task Materials

The rules employed were discussed in Chapt 6:

(i) If I finish my work, then I'll be home in time for dinner.
(ii) If the boss wants to see me, then I won't be home in time for dinner.
(iii) If the boss doesn't want to see me, then I'll be home in time for dinner.
(iv) If I don't finish my work, then I won't be home in time for dinner.

The 3 × 2 array of stimulus cards embodied all possible antecedents and consequents and their respective negates. The negations used in the rules and on the cards were in their morphologically abbreviated forms. Each subject received all four rules. Given the nature of the task materials, varying lexical content was not appropriate. The subjects' task was to select

a pair of cards from the array which either verified or falsified the rule. Since the materials were not homogeneous between antecedent and consequent no explicit left/right marking was necessary. If a subject placed the *I don't finish my work* card next to the *I don't make it home in time for dinner* card to verify, then this is equivalent to constructing the TT logical case ($\neg p, \neg q$ card case) to make the rule in (iv) true.

Procedure

Each subject was presented with a 3×2 array of the single sided thematic stimulus cards. Apart from the following typed instructions being given to the subject, the procedure was the same as for Experiment 1.

(7.3) You will be presented with a series of rules which concern the events described on the cards in the array before you. Your task will be to select two of the events described on the cards in the array which either make a given rule true, or make a given rule false, according to the instruction. If you have any questions please ask them now and not after you have started on the problems.

Results

Exactly the same series of analyses were performed for this experiment as for Experiment 1.

(i) Initial and Total Response Analysis

Table 7.8 shows the frequency with which subjects gave each truth table case on their initial verification and initial falsification of each rule. The TT construction was universally chosen initially as the only verifying case. The initial choice of the FT construction as a falsifier for the *if* $\neg p, q$ rule form in Evans (1972) and experiment 1 has completely disappeared in the present experiment. It was this initial choice which was observed to be responsible for the confirmation of Evans matching bias predictions. The analogous observation for the *if* $\neg p, \neg q$ rule of FF as the modal falsifier in the initial responses (.cf experiment 1) was also not observed. As would be expected, the matching bias prediction that there would be fewer TF falsifiers constructed for rules with negative antecedents than affirmative antecedents was not replicated. Neither was the prediction that there would be

Table 7.8

The frequency with which subjects gave each truth table case on their initial verification and initial falsification of each rule in the Thematic Construction Task.

Rule	Verification				Falsification			
	TT	TF	FT	FF	TT	TF	FT	FF
(i) If p, then q	24	0	0	0	0	18	2	4
(ii) If p, then ¬q	24	0	0	0	0	18	2	4
(iii) If ¬p, then q	24	0	0	0	0	17	4	3
(iv) If ¬p, then ¬q	24	0	0	0	0	15	6	3

more TF falsifiers constructed for rules with negative consequents. Along with the observations made in experiment 1 concerning the lack of matching and the failure of a misinterpretation hypothesis to capture the anomalous *if* ¬*p*, ¬*q* rule interpretation, this result offers strong support for a negative processing deficit account of the abstract results. Both the anomalous initial falsifiers were only observed in the abstract version of the task where subjects must compute the relevant contrast classes prior to establishing the appropriate class inclusion relation.

Table 7.9 shows the total frequency with which each subject gave each truth table case on verifying and falsifying each rule. In subject's total responses, neither matching prediction was borne out. The main expected difference between rule interpretations was that there should be more FF verifiers and more FT falsifiers for rules with negative antecedents than affirmative antecedents. Both expectations where borne out but the FT instance was picked as falsifying only marginally more often for negative antecedent rules than affirmative antecedent rules. However, the result for the FF instance was clear cut. There were significantly more FF verifiers for negative antecedent rules than affirmative antecedent

Table 7.9

The total frequency with which each subject gave each truth table case on verifying and falsifying each rule for the Thematic Construction Task

Rule	Verification				Falsification			
	TT	TF	FT	FF	TT	TF	FT	FF
If p, then q	24	0	1	9	0	21	11	4
If p, then ¬q	24	0	0	9	0	20	9	4
if ¬p, then q	24	0	0	13	0	20	14	3
if ¬p, then ¬q	24	0	0	18	0	20	12	3

rules ($p < 0.005$, one tailed). This belies a generalised material implication to material equivalence switch between abstract and thematic versions (cf. section (iv) "Between tasks analysis").

(ii) Psychological truth tables

Table 7.10 shows the result of plotting the frequency of true, false and "irrelevant" (non-constructed) classifications of each truth table case. Consistent with Evans (1972), the TT case was significantly more often constructed as true than false for all rules ($p < 0.0005$, one-tailed Binomial test). The TF case was significantly more often constructed as false than true for all rules ($p < 0.0005$, one tailed Binomial test). As in Experiment 1, differences between all values are reported, including "irrelevant". All tests are 1-tailed Binomial tests where the direction is predicted by the defective truth table/taxonomic constraint interpretation and two-tailed otherwise.

For the *if p,q* rule form, FT was constructed significantly more often as false than true ($p = 0.006$, 2-tailed). It was treated as irrelevant more often than false, although not significantly

Table 7.10

The frequency of true, false and "irrelevant" (non-constructed) classifications of each truth table case for the Thematic Construction Task (N = 24; modal response is in italics).

Rule	Truth Value	Truth Table Case			
		TT	TF	FT	FF
(i) If p, then q	T	*24*	0	1	9
	F	0	*21*	11	4
	?	0	3	*12*	*11*
(ii) If p, then ¬q	T	*24*	0	1	8
	F	0	*20*	9	4
	?	0	4	*14*	*12*
(iii) If ¬p, then q	T	*24*	0	0	*13*
	F	0	*20*	*14*	3
	?	0	4	10	8
(iv) If ¬p, then ¬q	T	*24*	0	0	*18*
	F	0	*20*	*12*	3
	?	0	4	12	3

0.006, 2-tailed). It was treated as irrelevant more often than false, although not significantly (p = 0.5, 1-tailed), and more often than true (p = 0.002, 1-tailed) . FF was constructed more often as true than false, although not significantly (p = 0.266, 2-tailed). It was treated as irrelevant more often than true, albeit not significantly (p = 0.332, 1-tailed), and more often than false but not significantly (p = 0.059, 1-tailed). For the *if* p, $\neg q$ rule form, FT was significantly more often constructed as false than true (p = 0.022, 2-tailed). It was more often treated as irrelevant than false, although not significantly (p = 0.202), and than true (p = 0.0005, 1-tailed). FF was more often constructed as true than false, but not significantly (p = 0.388, 2-tailed). It was treated as irrelevant more often than true, albeit not significantly (p = 0.252, 1-tailed), and more often than false (p = 0.038, 1-tailed). For the *if* $\neg p$, q rule form, FT was constructed significantly more often as false than true (p < 0.0005, 1-tailed), and more often than it was treated as irrelevant, although not significantly (p = 0.271, 1-tailed). FF was constructed as true significantly more often than false (p = 0.011, 1-tailed), and more often than it was treated as irrelevant, albeit not significantly (p = 0.192, 1-tailed). For the *if* $\neg p$, $\neg q$ rule form, FT was constructed as false significantly more often than true (p = 0.0005, 1-tailed), but it was treated as irrelevant in equal numbers (see below for the justification of treating this as false). FF was constructed significantly more often as true than either false (p = 0.001, 1-tailed) or irrelevant (p < 0.001, 1-tailed).

From table 7.10, the three valued truth tables in table 7.11 were extracted. 9 out of 16 (56.25%) of these truth table entries reflect significant differences from both remaining possible response categories. This is a far less clear cut result than the abstract version. However, all the differences in modal responses between the abstract and thematic versions were significant (cf. section (iv): "Between tasks analysis").

Although there were equal numbers of irrelevant and false classifications for the FT

Table 7.11

Psychological truth tables extracted for each rule form for the Thematic Construction Task.

		Rule-Form							
		if p, q		*if p, ¬q*		*if ¬p, q*		*if ¬p, ¬q*	
		q		q		q		q	
		0	1	0	1	0	1	0	1
p	0	?/1/0	?/0	?/0	?/0	0	1	1	0
	1	0	1	1	0	1/?	0/?	0/?	1

instance in the *if* $\neg p$, $\neg q$ rule form it has been treated as falsifying in these truth tables. There are two reasons for this. First, the irrelevant category is a response by omission, rather than by commission. Where there is an ambiguity, therefore, it should be resolved in favour of the subjects positive responses. Second, the rationale behind these experiments is to extract the general pattern in subjects responses, rather than looking at individuals' global response patterns, many of which are uninterpretable. However, in order to resolve the ambiguity, it can be noted that if FT was being treated as irrelevant then a response profile where TT and FF were constructed as verifying and only TF as falsifying would be expected to predominate in subjects global responses. 29% of subjects global responses conformed to this pattern. However, 46% of subjects responses accorded with a response profile where TT and FF were constructed as verifying and where TF *and* FT were constructed as falsifying. This accords with the hypothesis that FT is correctly classified as false for this rule.

(iii) Defective truth tables?

As for Experiment 1 the same analyses as Evans (1982) were carried out for the present data. It was argued that due to the interpretational changes between abstract and thematic tasks the response profile for the thematic version would be expected to differ from the monotonic trend observed in the abstract task.

Table 7.12 pools the results over the four rules classified by (a) logical case and (b) matching case. From table 7.12 it can be seen that in the present data the defective truth table account was not directly corroborated. 48% of subjects FT responses indicated falsity and 50% of their FF responses indicated true. However, this would be expected given *half* the rules were being treated as equivalence. So, this does not undermine the view that false antecedents are irrelevant *on affirmative antecedent rules* when an enablement relation is being implicitly asserted to exist between antecedent and consequent. Although this is a sound interpretation of the table, it should be recalled that in the psychological truth tables, the even split between affirmative and negative antecedent rules is not observed. Nonetheless, the overall reduction in irrelevants is precisely in line with the predicted interpretational changes. For matching case, the *trend* for more irrelevants was not observed. Mismatching cases yielded only minimally more irrelevants. Moreover, where they were expected to occur was where they were least observed. A sharp rise would be expected between the 0-mismatching case and 1-mismatch cases and between single mismatches and

- double mismatches ?

Table 7.12

Results pooled over the four rules and classified by (a) Logical Case and (b) by Matching Case. Results are percentage frequencies (N = 24).

Case	Classification		
	True	False	Irrelevant
(a) Logical			
TT	*100*	0	0
TF	0	*84*	16
FT	2	48	*50*
FF	*50*	15	35
(b) Matching			
p, q	44	39	17
p, ¬q	39	38	23
¬p, q	34	36	30
¬p, ¬q	35	34	31

clearly demonstrated by looking to the frequency of irrelevants by mismatching case within rules. Table 7.13 shows the frequency of irrelevants by rule type as a function of mismatching case. Few of the expected increases in irrelevants predicted by matching occur within rules and where they do occur they are marginal, and in many cases go in the opposite direction. The failure to observe any non-marginal differences between 1 and 2 mismatches questions the viability of the matching hypothesis.

Table 7.14 shows the result of plotting the frequency of true, false and irrelevant responses for rule forms/logical cases which constitute a double match, as in Evans (1982). Table 7.14 fails to replicate Evans (1982/1972). The FF instance for the *if ¬p, ¬q* rule has

Table 7.13

The frequency of "irrelevants" (non-constructed) classifications for each rule and mismatching case (Thematic).

Rule	Number of Mismatches			
	0	1	1	2
(i) If p, then q	0	3	12	11
(ii) If p, then ¬q	4	0	12	14
(iii) If ¬p, then q	10	8	1	4
(iv) If ¬p, then ¬q	3	4	12	0
	17	15	37	29

Table 7.14

The percentage frequency of construction of each logical case on rules where they constitute a double match, pq. (N = 24).

Logical Case	Rule	% Frequency True	False	Irrelevant
TT	If p, then q	100	0	0
TF	If p, then ¬q	0	83	17
FT	If ¬p, then q	0	58	42
FF	If ¬p, then ¬q	75	12.5	12.5

7.14 fails to replicate Evans (1982/1972). The FF instance for the *if* $\neg p$, $\neg q$ rule has switched to true. The high number of irrelevants for the *if* $\neg p$, q rule form undermines the the matching hypothesis which argues that double match cases should alleviate response suppression. Moreover, the fact that for the *if* $\neg p$, $\neg q$ rule, precisely the reverse is observed suggests strongly that matching is not determining subjects responses.

(iv) Evans' matching indices

Antecedent and consequent matching indices were calculated for each subject (Evans, 1983b). AMI had a mean of 0.75 and a standard deviation of 1.09, CMI had a mean of 0.417 and a standard deviation of 1.151. If an index is positive then this indicates the presence of matching, by the criterion laid down by Evans (1983b). A significant majority of subjects produced positive indices on AMI (17 +, 3 -, $p < 0.006$, 1-tailed Binomial test) but not on CMI (15 +, 5 -, $p = 0.151$).

Evans (1983b) provides the means and standard deviations of the matching indices he obtained which permits a comparison. However, each of his subjects performed the task twice, so to obtain comparable statistics the means and the standard deviation were doubled. AMI was higher than observed in Evans' (1983b) E-group, but not significantly. However, CMI was significantly lower in the present experiment (unequal Ns, independent samples t-test, $t = 2.238$, 62 df., $p < 0.05$, 2-tailed).

As in experiment 1, these results seem to support Evans contention that there is a significant matching bias observed in these tasks. So pure matching scores were again computed which correct for the interpretational effects predicted by pragmatic context theory. There has been a change in the predicted interpretations such that for negative antecedent

rules, no irrelevants are predicted. Therefore, for these rules the occurrence of any irrelevant response must be assigned to the pure matching category.

The results by rule form were as follows. For the *if p, q* rule form although a majority of subjects were non-matchers the result of a Binomial test was not significant (14 0, 10 +, $p = 0.271$; PMS: M = 0.417, SD = 0.493). There was a significant majority of non-matchers for the *if p,* $\neg q$ rule form (17 0, 7 +, $p = 0.032$; PMS: M = 0.458, SD = 0.763). For the *if* $\neg p$, *q* rule form the majority of subjects were pure matchers but not significantly (9 0, 15 +, $p = 0.154$; PMS: M = 0.917, SD = 0.812). For the *if* $\neg p$, $\neg q$ rule form, although there were more pure matchers than non-matchers, the result of the Binomial test was not significant (11 0, 13 +, $p = 0.419$; PMS: M = 0.792, SD = 0.815).

There was less pure matching for the *if p, q* than the *if p,* $\neg q$ rule form (comparing means), but not significantly ($t = 0.272$, n.s). However, there was significantly less pure matching in the *if p, q* rule form than both negative antecedent rule forms: *if* $\neg p$, *q* ($t = 2.627$, $p < 0.02$) and *if* $\neg p$, $\neg q$ ($t = 2.386$, $p < 0.05$). There was significantly less pure matching for the *if p,* $\neg q$ rule form than for the *if* $\neg p$, *q* rule ($t = 2.110$, $p < 0.05$) and there was less pure matching on the *if p,* $\neg q$ than the *if* $\neg p$, $\neg q$ rule form, but not significantly ($t = 1.360$, n.s.). There was no significant difference in the occurrence of pure matching between the negative antecedent rules ($t = 0.720$, n.s.).

There was significantly less pure matching on affirmative antecedent rules than for negative antecedent rules. This belies any general matching strategy as an explanation of these results. PMS scores were significantly higher for negative antecedent rules. This seems to argue that despite the appropriate interpretations being adopted, subjects still incur some processing deficit for antecedent negations when a non-taxonomic interpretation is appropriate. However, this result could not have been predicted on the basis of the consideration that explicit negations were included on the cards in the array (Evans, 1983b). This manipulation would be expected to alleviate response suppression across the board, and not selectively between rules. However, whether the results on the thematic task can be put down to an effect of a simple perceptual matching strategy or a release from response suppression can only be determined via the between tasks analysis.

(v) Between tasks analysis: construction tasks

It was argued in the last chapter that if the anomalous results on negative antecedent rules for the abstract version were due to a non-taxonomic constraint misinterpretation, because of their disjoint expression, then similar initial response profiles would be expected between the abstract and thematic versions. However, it has already been observed, qualitatively, that this is not the case. This was also confirmed in the quantitative comparisons. For the *if* $\neg p$, q rule form, the TF instance was constructed initially significantly less often in the abstract task than the thematic task (McNemar tests, 2-tailed, $p < 0.001$) and the FT instance was constructed significantly more often in the abstract task ($p < 0.001$). For the *if* $\neg p$, $\neg q$ rule form, less subjects constructed the TF instance as the initial falsifier for the abstract task than the thematic task, albeit not significantly. However, significantly more subjects constructed the FF instance as their initial falsifier in the abstract task than the thematic task ($p < 0.01$). These results argue strongly against the misinterpretation hypothesis and for a negatives processing deficit account of the anomalous results. This hypothesis is consistent with the hypothesis that misinterpretations are implicated in selection tasks (cf. chapter 5), since in those tasks an information gaining eduction is also required. Concern centres not on truth conditions but on the relations which govern the rules information gaining behaviour.

In discussing Evans (1983b) it was suggested that a possible cause of the significant reduction in the matching indices for the E-group was a shift to an even more clear cut adoption of a biconditional interpretation for all the rule forms. It was argued that when explicit negations are present on the cards this could lead subjects to adopt a simple perceptual matching strategy. Such a strategy would predict a generalised shift towards a biconditional interpretation from the abstract to the thematic task. More FF verifiers and more FT falsifiers would be expected for all rules. The shifts in Evans (1983b) data would be expected to be small since even the I-group subjects seemed to be predominantly treating the conditional as equivalence. This was only the case for the *if* $\neg p$, q rule form in the present abstract experiment, so if a perceptual matching strategy were being adopted, then significant increases in FF verifiers and FT falsifiers would be expected for all rules.

Qualitatively, it can be observed that there has been a general interpretational shift towards equivalence in the thematic version. For all rules more FT instances were constructed as false in the thematic variant than in the abstract variant and more FF instances were treated as true. However, it was only for negative antecedent rules that this shift resulted in changes in modal responses. Although the shift occurred for affirmative antecedent rules it

was not sufficiently pronounced to have this effect. With this in mind, the results of the quantitative comparisons were as follows. The McNemar test was used throughout, but where the sum of the change cell frequencies dropped below 10 (ie. expected cell frequency < 5), the Binomial test was used (cf. Siegel & Castellan, 1988:79). For the *if p, q* rule form there were significantly more FF verifiers in the thematic version than the abstract version ($p = 0.02$, 1-tailed Binomial test). However, although there were more FT falsifiers in the thematic version this was not significant, as assessed by the Binomial test. For the *if p, ¬q* rule form there were again significantly more FF verifiers in the thematic version than the abstract version ($p = 0.035$, 1-tailed Binomial test). However, although there were more FT falsifiers in the thematic version this was not significant, as assessed by the Binomial test. For the *if ¬p, q* rule form, a highly significant increase in FF verifiers for the thematic version over the abstract version was observed ($p < 0.005$, McNemar test). However, there was a significant *reduction* in FT falsifiers for the thematic version ($p = 0.02$, 1-tailed, Binomial test). For the *if ¬p, ¬q* rule form a highly significant increase in FF verifiers was observed for the thematic version over the abstract version ($p < 0.0005$, McNemar test), a similar increase was observed for FT falsifiers but not as significantly ($p < 0.05$, McNemar test).

There has been a general shift towards an equivalence interpretation. However, it is far more pronounced for the negative antecedent rules where it was predicted to occur on the basis of pragmatic context theory. Only two significant increases were observed for affirmative antecedent rules. FF was constructed as a verifier more often in the thematic version than the abstract version for both rules, however neither represented the modal response for these rules and instances, so there has been no overall shift for the affirmative rules. This contrasts with the negative antecedent rules where the increases were highly significant for all instances in the predicted direction and each constituted the modal response for the instance FF or FT. All, that is, except for the FT falsifier in the *if ¬p, q* rule form. Although this was the modal response in the thematic version, there were significantly more of these responses in the abstract version. This is solely due to the anomalous result in the abstract case which is under investigation. However, the direction of this result is inconsistent with a general shift across all rules as the result of a perceptual matching strategy, as is the whole pattern of these comparisons. If subjects were adopting this strategy, then the effects would be expected across the board and not selectively between rules. However, some perceptual matching may be present and could account for the increases on affirmative antecedent rules.

The tests conducted for negative antecedent rules also function to confirm the hypothesis of

interpretational changes as a function of whether a taxonomic or non-taxonomic constraint is being described in the rule. However, although each result represented a shift in modal response it could still be the case that a significant reduction in the abstract modal response was not observed. Hence three further McNemar tests were conducted to ensure that the significant increases observed in the thematic results were accompanied by significant reductions in the modal responses observed in the abstract task. There were significantly less FF irrelevants in the thematic version than in the abstract version for the *if* $\neg p,q$ rule form ($p < 0.005$, McNemar test). For the *if* $\neg p$, $\neg q$ rule form, significantly less FT irrelevants ($p < 0.05$) and significantly less FF falsifiers ($p < 0.005$) were observed in the thematic variant. Taken together these results are strongly supportive of an interpretational change between the abstract and thematic variants in accordance with the predictions of pragmatic context theory. Whether the rule expresses a taxonomic or non-taxonomic constraint significantly affects subjects distribution of true/false/irrelevant constructions and in the predicted directions.

The misinterpretation hypothesis has been discounted as responsible for the anomalous abstract results and a simple perceptual matching strategy rejected as determining subjects changes in interpretations. Moreover, in both experiment 1 and experiment 2 there was a patent absence of results which would support the possibility that the explicit negations are simply alleviating response suppression in accordance with Evans (1983b). However, the possibility was tested for by making the same comparisons between matching indices carried out by Evans (1983b). The AMI was significantly lower for the thematic version than the abstract version ($t = 2.64$, $p < 0.01$) in accordance with the release from response suppression hypothesis. However, although the CMI was lower on the thematic variant, it was not significantly lower ($t = 0.826$, n.s.).

Using Evans' (1983b) indices yields an equivocal result in the present data. However, the demonstrable presence of highly differentiated interpretational effects in these experiments renders the use of Evans' indices invalid since they are computed over the whole response record. Hence differences were tested for using the PMSs for each rule form. For the *if* p, q rule form, the PMS was *larger* for the thematic version than the abstract version, although not significantly ($t = 1.165$, n.s.). For the *if* p, $\neg q$ rule form the PMS was again *larger* for the thematic version than the abstract version, this time significantly in a 1-tailed test ($t = 2$, $p < 0.05$). This was in the opposite direction to that predicted by the release from response suppression hypothesis. For both negative antecedent rules the PMSs were lower for the thematic version but not significantly, *if* $\neg p$, q ($t = 1.238$, n.s.) and *if* $\neg p$, $\neg q$ ($t = 1.100$, n.s.).

These results argue against the position that the inclusion of explicit negations functions to override an early processing linguistic heuristic. The data has clearly differentiated between negative antecedent and affirmative antecedent rules. The lower PMSs for the thematic version on negative antecedent rules seems to reflect the less equivocal interpretation available when appropriate thematic content is included. And the higher PMSs for affirmative antecedent rules seems to reflect the possibility that some pure matching is occurring. In general, the vast majority of the results obtained in these experiments are consistent with the interpretational effects predicted by pragmatic context theory. Only a small minority of the matching predictions were confirmed and where they were it was due to a failure to correct for other possible causes of irrelevant responding.

Discussion

These results offer strong support for the hypothesis of an interpretational effect determined by the nature of the relations asserted to exist between antecedent and consequent of a conditional and the way these relations interact with negation. The predictions of pragmatic context theory have been broadly confirmed for the construction tasks. The differences observed confirm that negation interacts in a different manner given different relations, ie. here class inclusion versus enablement/impediment. It appears this interaction differentially affects peoples ability to process negative information.

There are three possible sources of the apparent discrepancy between Evans (1983b) and the present results. First, these experiments were construction not evaluation tasks. In an evaluation task a complete instances is presented to a subject in isolation from other possibilities. However, in a construction task the domain over which the implicit universal quantifier varies is before the subject in the card array, presenting the whole range of possibilities and crucially all the possible members of the relevant contrast classes. Moreover, having to construct an instance rather than evaluate it in isolation, will involve engaging higher level processes: the correct selection has to be drawn from a range of possibilities. Second, in the present experiment the negations were present on the cards in their morphologically abbreviated form in the thematic version, they were not explicit. However, they were also in this form in the rules. It is hard to assess any consequences of this change since in natural language the need to employ an explicit negation is minimal: natural language is replete with opposites, antonyms, dimensions of variation etc. Because the task materials mirrored the example provided by Evans (1983b) meant that the most felicitous expression was using the morphologically abbreviated form. Third, the possibility of

population variations must always be admitted. It has already been observed that Evans' (1983b) subjects predominantly treated the rules as biconditionals. Moreover, the CMI index was significantly lower in the present data for both abstract and thematic tasks than in Evans' (1983b). Perhaps Florida University students are more emphatic than Edinburgh University students, and more unwilling to believe the context may change after all and prove them wrong.

The anomalous results obtained on the abstract version would appear to be unequivocally attributable to a processing deficit incurred due to the presence of negative components. Moreover, the specific locus of the processing deficit appears to lie with antecedent negation when a class inclusion constitutes the appropriate interpretation of the relation described by the conditional. Pragmatic context theory argues that in the right contexts negations serve to identify contrast classes. Given a negated constituent subjects must actively construct the contrast class either by accessing prior knowledge concerning the taxonomies and relations in which that constituent participates or by direct appeal to the environmental context. The emphasis is on the fact that the contrast class needs to be *constructed.* In the constructional process subjects must begin with the *affirmative constituent* and use it to access the appropriate contrast class. If the process proves difficult for any of a variety of reasons, this may cause subjects to abandon the constructive process and simply *match* the affirmative constituent. Hence a constructional view of negation may not only explain the apparent processing deficit but also the albeit limited occurrence of matching in these experiments. A similar constructional complexity hypothesis underlies Johnson-Laird and Steedman's (1978; and Johnson-Laird, 1983) account of errors in syllogistic reasoning.

It has often been observed that the *abstract* tasks are in fact more *concrete* than thematic tasks. In the latter subjects can integrate their interpretation of the rules with the appropriate semantic structures and relevant prior experience. However, it seems that,

(7.4) "...the abstract tasks involve the mental manipulation of concrete objects - in the absence of - the semantic structures and relevant experience for understanding and relating the sentences as structural wholes." (Evans & Lynch, 1973:396)

Subjects responses in the abstract construction task may be explained by allowing that the concrete presence of the array provides the environmental circumstances relative to which subjects attempt to construct the relevant contrast classes and subsequently establish the relationship between antecedent and consequent.

A series of diagrams will be used, called Aberrant Constructive Venn diagrams (ACVs), to

reconstruct the processes involved in subjects reasoning. This has a different aim from Johnson-Laird's theory of mental models. He was concerned to provide a model which allowed logically pristine performance, but which admitted graded constructional complexity to combine with resource limitations to explain observable error patterns. In the present case the principle observation which coheres with rationality is that subjects interpret negation in these contexts constructively. However, the way these processes interact with the concrete context provided by the card array may lead to systematic errors due to the process itself rather than resource limitation. Hence, the following account has far more limited aims. It does not constitute a generalisable model of conditional reasoning. Rather it provides a task specific model of how the assertion of a particular relation, class inclusion, interacts with negation in a particular environmental context.

The assertion of a class inclusion relation is counterintuitive in the context of the task. Treating the relation as class inclusion requires subjects to abstract away to card *pairs*. As Wason & Green (1984) have observed, in the standard selection task the relevant unifying objects are the cards. But because the rule focuses attention on the card sides subjects may treat the rule as *disjoint*. In which case class inclusion would be an inappropriate interpretation of the implicit relation. Similar arguments apply here, but even more so, as the card *pairs* do not actually form a concrete unified object. So, to get the correct interpretation, a rule like:

(7.5) If there is a blue circle on the left, there is not a green diamond on the right.

has to be interpreted as:

(7.6) The set of card pairs with a blue circle on the left is included in the set of cards pairs where the card on the right has any colour/shape combination in the array other than a green diamond.

The convoluted nature of this description serves to indicate the complexity of the task subjects confront. It is surprising, therefore, that apart from one or two processing errors, subjects manage rather well.

Figures 7.1 to 7.4 show the arrays and ACVs for all rule forms. The constructive process subjects' must engage in to construct the contrast class is shown by the arrows indicating the process unfolding. The inclusion of the array is to show how its concrete presence is interacting with the interpretative process in order to construct the relevant contrast class and subsequently establish the appropriate relation of class inclusion between them. The array show the various cards: blue circle (BC), red diamond (RD), yellow square (YC),

Rule: If red circle left, green diamond right

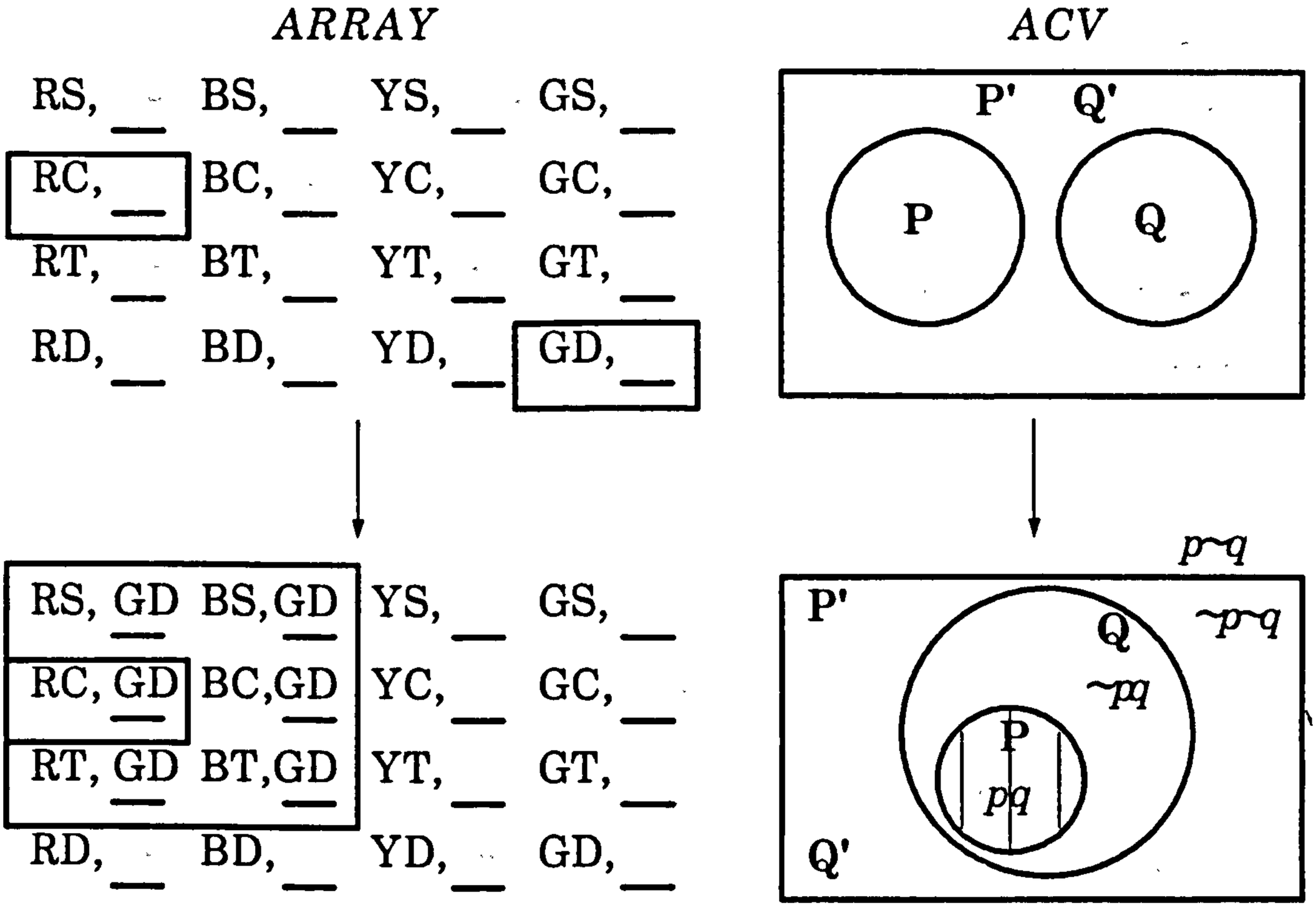

Figure 7.1 : ACV and Array for the if p, q rule form

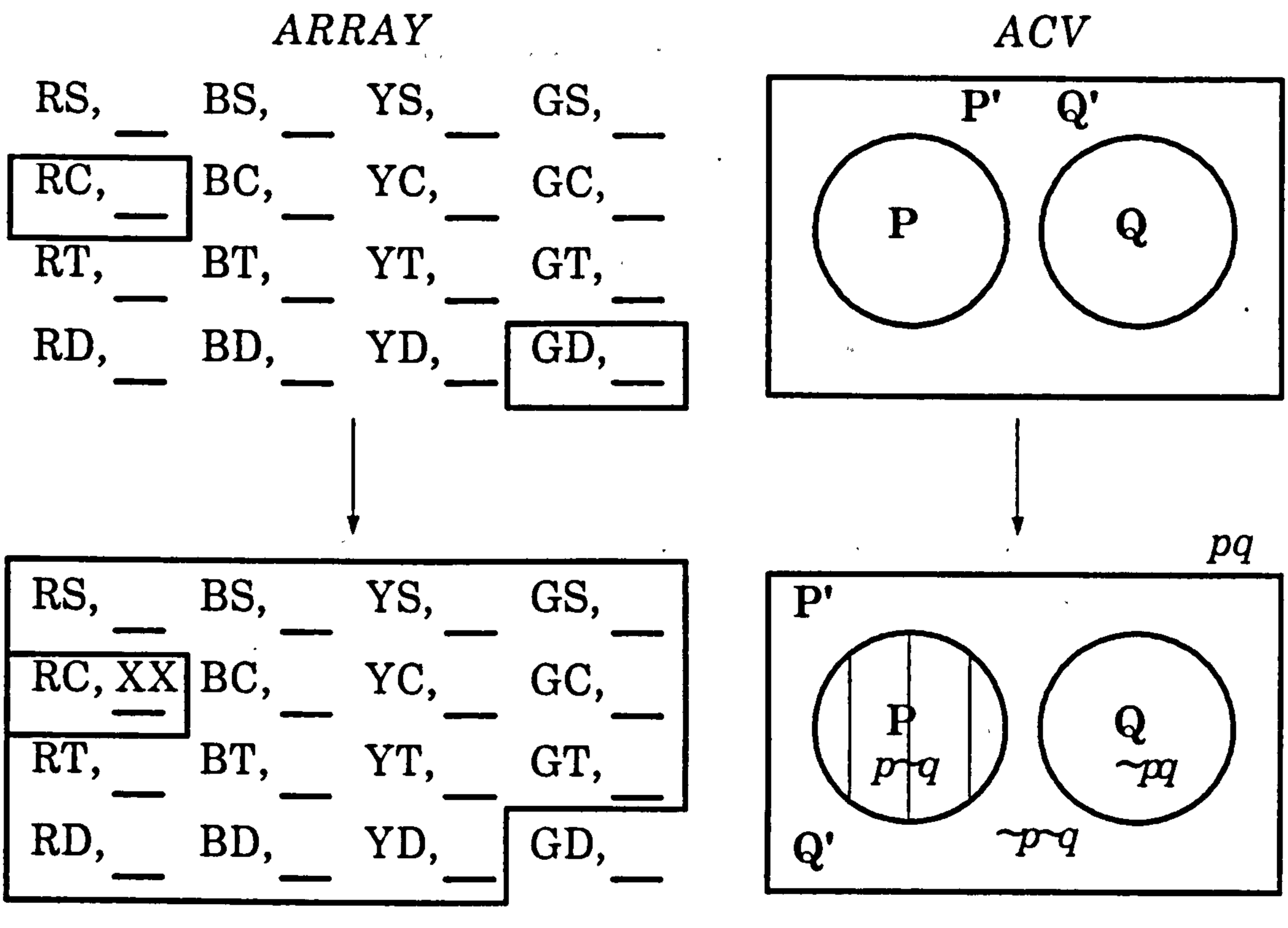

Figure 7.2: ACV and Array for the if p, ~q rule form

green triangle (GT) etc. The slots beside each coloured shape indicates that the relation is asserted between card pairs. Subjects are therefore attempting to identify contrast classes and pair up cards appropriately. The fact that the slots are to the right of the coloured shape, should not be confused with the left/right instructions in the rules. The ACV serves to provide a more perspicuous account of why subjects construct the various true, false instances, and why they don't construct others.

If p, q: (Figure 7.1) Subjects always begin by identifying the named items in the rules. This places two distinct set circles in the ACV. They then establish the relation between the named items by matching the consequent card to the slot in the antecedent card. The rule does preclude green diamonds being paired with other cards, nor other cards than

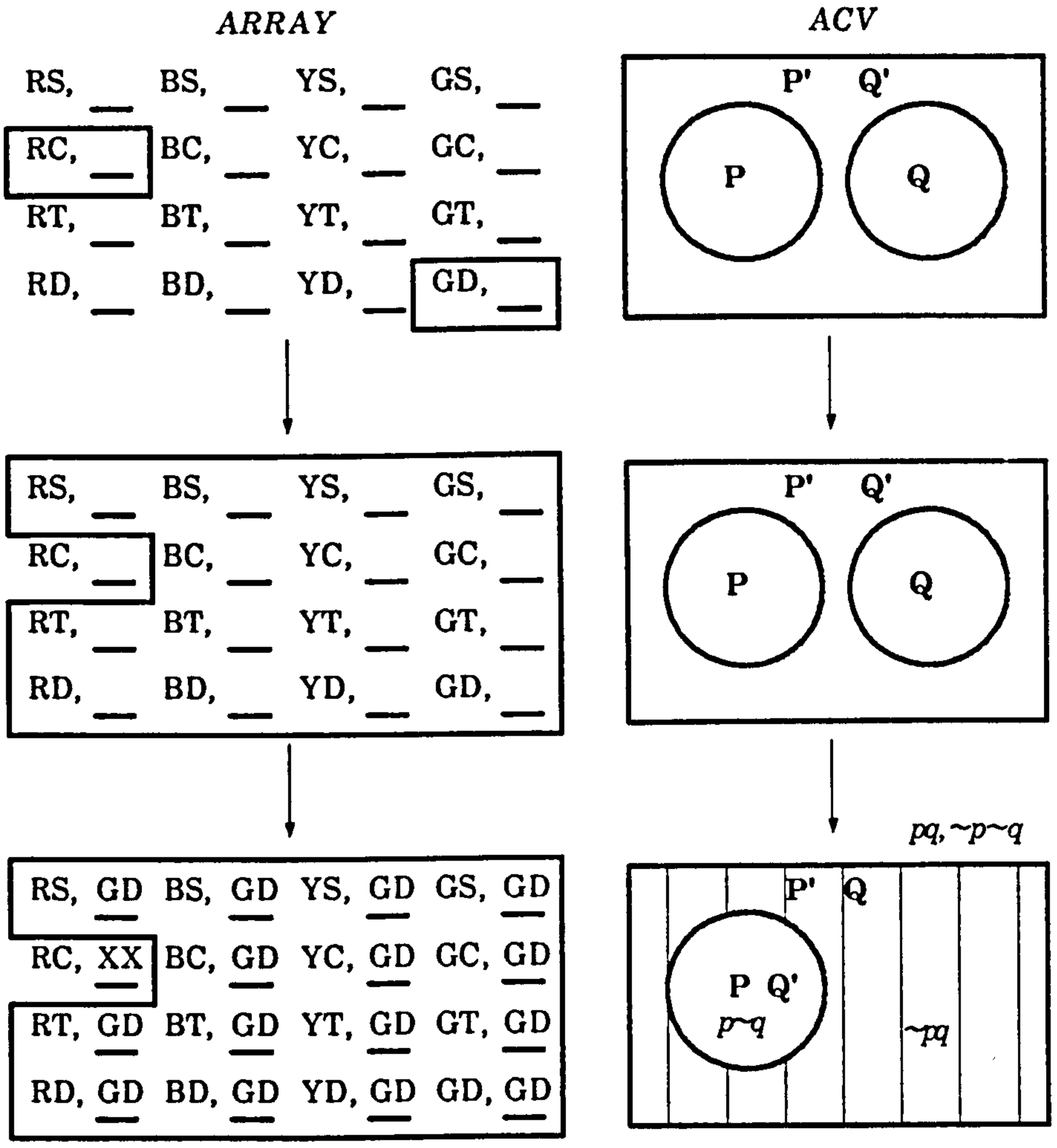

Figure 7.3: ACV and Array for the if ~ p, q rule form

either those explicitly mentioned being paired. The ACV, indicates that once these pairings have been made then a p, $\neg q$ pair can not be constructed, and hence is false. The other pairings do not involve the p card and are therefore not constructed and hence are treated as irrelevant. The cross hatched area indicates the true assignments.

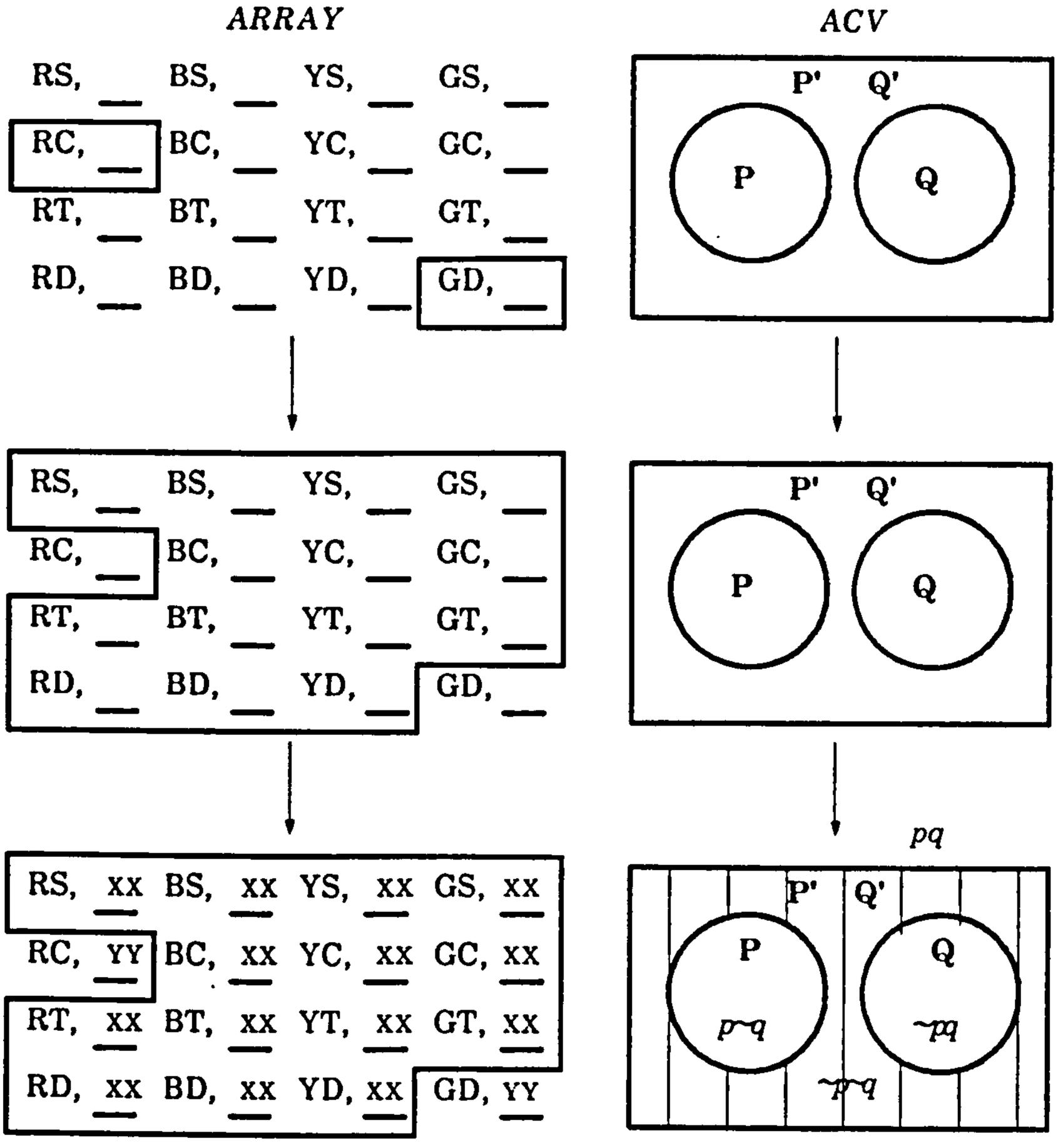

Figure 7.4: ACV and Array for the if ~ p, ~q rule form

If p, ¬q: (Figure 7.2) Again the subjects begin with the named values. They must then construct the contrast class for the consequent. In a normal Venn diagram representation this would involve simply switching complement markers, and the same basic diagram as in Fig. 7.1 would result. However, the concrete presence of the array drives subjects' constructive processes. And in the array the space of pairs which represents *not green diamond*

is the larger box in the lower array. *p* (RC) is in this space, and the appropriate relation is established by allowing any member within the larger box to be paired with RC. The XX functions as a variable in the diagram indicating that subjects believe anything in the larger box can be paired with RC. But equally items in the larger box minus RC can be paired with each other. This gives the pattern of true, false and irrelevants in the lower ACV. Only *pq* can not be constructed.

If ¬p, q (Figure 7.3) Again subjects begin by identifying the named items. They must then construct the contrast class for the *p* card (RC), this is shown by the large box in the middle array. Subjects then proceed to pair items. They take the rule to state that every item that is in the large box, ¬*p*, must be paired with *q*, the green diamond, which they proceed to do. But this means that the space of *q*s is now the same as the ¬*p*s. Subjects therefore assume that the only pairings for ¬*q* must be with *p* (RC), this is again indicated using the XX variable. But once the space is divided up in this fashion then neither *p*, *q* nor ¬*p*, ¬*q* pairs can be constructed.

If ¬p, ¬q: (Figure 7.4) Once the named items have been located, the joint contrast classes must be identified. Subjects will probably do this in the antecedent-consequent order, but for reasons of space the constructive process for both is embodied in the middle array. The large box indicates the co-incidence of contrast classes. Each item in the large box can be paired with any other item in the large box, indicated by the variable XX. *p* can be paired with any item in the large box as can *q*, indicated by the variable YY. Once the space has been divided up in this way, then only the *p*, *q* pair can not be constructed, it is therefore treated as false.

In conclusion, subjects error patterns are driven by the circumstantial features provided by the array and the requirement to construct the relevant contrast classes. When thematic materials are used, which require subjects to apply a non-taxonomic interpretation of the linguistic expressions on the cards then the systematic error patterns observed for the abstract version disappear: subjects no longer have to exploit their environment in interpreting the negations.

7.4 Experiment 3: Abstract Selection Task

Introduction

This experiment was an abstract selection task (Evans & Lynch, 1973). The complete introduction is given in Chapt 6.

Method

Subjects

The same 24 undergraduate psychology students from the University of Edinburgh served as subjects on this experiment as in Experiments 1 and 2. All subjects were tested individually on all conditions.

Design

Subjects were required to determine the truth or falsity of four conditional rules of the form: *if p, q; if p, ¬q; if ¬p, q; if ¬p, ¬q,* as in Evans and Lynch (1973). Each subject was given a different one of the 4! permutations of presentation order of the rules.

Task Materials

Cards with various coloured shapes were used in the present experiment. A different pair of shapes and pair of colours was used for each of the four rules. Thus, the lexical materials were different for each rule form. The lexical material was varied randomly between rule forms. Each card had one shape on one side and the other shape on the other side; taking all colour-shape combinations yielded the four cards used for each rule. These materials were analogous to Wason's (1969) form group, where only two colours and shapes were used and subjects knew there was shape of each type on each card. Since Wason (1969) simply replicated the standard result (in his initial choice group) using the familiar, A, K, 2, 7 materials, if any of the predicted effects are observed this can only be attributed to the binary situation interacting with the context of the negations paradigm. An example rule is

shown below:

(7.7) If there is not a yellow square on one side, then there is a blue triangle on the other side.

Procedure

Each subject was given four of the double sided colours and shapes stimulus cards to look at and handle. Subjects were told to familiarise themselves with the cards but not to learn them. They were allowed to handle the cards for as long as they wanted (normally between 20 and 30 secs.). The cards were then placed in front of the subject in the appropriate orientation such that there was a p, a $\neg p$, a q and a $\neg q$ face uppermost, but randomly juxtaposed in a line. The subject was then given the following typed instructions and told to read them carefully:

(7.8) You will be presented with arrays of four cards like the ones you have just seen and handled, naturally only one side of each card will be visible. Only two shapes will be involved in each problem and it can be assumed that each card has one shape on one side and the other shape on the other side.
For each array, you will be given a rule which applies to certain combinations of the coloured shapes which appear on the cards.
Your task is to indicate which card or cards you would have to turn over in order to test whether the rule is true or false.
For each array and rule, I will simply ask you to point to the cards you think should be turned over; do not actually turn the cards over.
You may take as long as like over the problems, once you have finished one you will be presented with the next. In this task there are four problems in all.
If you have any questions, please ask them now and not after you have started the problems

Subjects were then told that they could keep the instructions to refer to. The four rules were then presented one at a time on a typed sheet in the order of presentation assigned and from the set of materials assigned.

Results

It was argued in Chapt. 6 that when the selection task data is analysed exhaustively there are three levels of comparative analysis. Along with the overall results and a Reich and Ruth (1982) style score, the results will be reported by levels.

(i) Overall Results

Table 7.15 shows the frequency with which subjects gave each logical case as being necessary to test for the truth or falsity of each rule. This table is in the standard format for presenting Evans negations paradigm data. It was argued in Chapt. 6 that simply presenting the data by logical case obscured the more detailed levels of analysis which can be performed for verification and falsification when the data are presented by card case. Table 7.16 shows the same data as table 7.15, but by card case.

(ii) Reich & Ruth (1982) Scoring

From Chapt. 6, it should be recalled that Reich & Ruth (1982) added the numbers of selections where each strategy makes unique predictions to obtain a score indicating the popularity of each strategy. These were as follows:

(1) Choice of $\neg q$ on *if* p, q rule form = falsifying,
(2) Choice of $\neg q$ on *if* p, $\neg q$ rule form = verifying,

Table 7.15

The frequency with which subjects gave each logical case as being necessary to test for the truth or falsity of each rule for the Abstract Selection task.

Rule	TA	FA	TC	FC
(i) If p, q	19	6	15	5
(ii) If p, ¬q	20	6	4	16
(iii) If ¬p, q	14	8	17	10
(iv) If ¬p, ¬q	21	6	13	10
Overall %	77	27	51	43

Table 7.16

The frequency with which subjects gave each card case as being necessary to test for the truth or falsity of each rule for the Abstract Selection task.

Rule	*p*	¬*p*	*q*	¬*q*
(i) If p, q	19	6	15	5
(ii) If p, ¬q	20	6	16	4
(iii) If ¬p, q	8 } 14	14	17	10
(iv) If ¬p, ¬q	6 }	21	10	13
Overall %	55	49	60	33

(3) Choice of p on *if* $\neg p$, q rule form = matching,
(4) Choice of $\neg q$ on *if* $\neg p$, q rule form = falsifying,
(5) Choice of p on *if* $\neg p$, $\neg q$ rule form = matching,
(6) Choice of $\neg q$ on *if* $\neg p$, $\neg q$ rule form = verifying.

The results of deriving similar scores for the present data indicated the following preferences:

(7.9). Verification (17) > Falsification (15) > Matching (14)

This result accords neither with a similar analysis of Evans & Lynch (1973):

(7.10) Falsification (18) = Matching (18) > Verification (9)

Nor with the analysis of Reich & Ruth (1982):

Abstract: Matching (27) > Verification (16) > Falsification (13)

Thematic: Verification (19) > Falsification (15) > Matching (4)

Reasons for this discrepancy are discussed below. These analyses presents a descriptive feel for the pattern of results. In the following analyses the data are subject to inferential statistical analysis to further determine the nature of subjects strategies in the present experiment.

(iii) Level I Analyses

• Matching

Keeping card case constant, matching predicts (i) more p card selections than $\neg p$ card selections ($p > \neg p$), and (ii) more q card selections than $\neg q$ card selections ($\neg q > q$). Although prediction (i) was borne out (p:53; $\neg p$:47) this was far from significant ($r = 8$, $N = 16$, n.s., all results are one tailed sign tests). Prediction (ii) was significantly confirmed ($r = 4$, $N = 17$, $p < 0.025$).

• Verification

Keeping logical case constant, verification predicts (i) more TA selections than FA selections (TA > FA), and (ii) more TC selections than FC selections (TC > FC). Prediction (i) was significantly confirmed ($r = 1$, $N = 21$, $p < 0.0005$), however, although prediction (ii) was borne out (TC:49; FC:41) this was not significant ($r = 6$, $N = 15$, n.s.).

• **Falsification**

Keeping logical case constant, falsification makes the same antecedent prediction as verification, but predicts the exact opposite for the consequent cases, ie. FC > TC. As for verification, prediction (i) was borne out, but prediction (ii) went in the opposite direction. This is expected, since verification and falsification make mutually exclusive predictions for the consequent case and the verification prediction was borne out.

At this level of analysis it would appear that subjects are verifying on the antecedent cards and matching on the consequent cards. Falsification appears ruled out because all consequent results go in the opposite direction to that predicted by this strategy.

(iv) Level II

• **Matching**

The four level II matching bias predictions have already been introduced. For the abstract version, the prediction that there would be more TA selections for rules with affirmative antecedents than negative antecedents was borne out (p:39; $\neg p$:35), but far from significantly (r = 4, N = 10, n.s.). There were more FA selections for rules with negative antecedents than for those with affirmative antecedents ($\neg p$:14; p:12); but again this result fell well short of significance (r = 7, N = 14, n.s.). The two consequent predictions were both significantly confirmed. There were more TC selections for rules with affirmative consequents than negative consequents (r = 3, N = 17, $p < 0.01$), and more FC selections for rules with negative consequents than affirmative consequents (r = 4, N = 16, $p < 0.05$).

• **Verification**

The prediction that there would be more p card selections for rules with affirmative antecedents than negative antecedents was significantly confirmed (r = 1, N = 18, $p < 0.0005$), as was the prediction that there would be more $\neg p$ card selections for rules with negative antecedents than affirmative antecedents (r = 1, N = 19, $p < 0.0005$). However, although both consequent predictions were borne out neither was significant: more q card selections on rules with affirmative consequents than those with negative consequents (q:32; $\neg q$:26; r = 4, N = 13, n.s.); more $\neg q$ card selections on rules with negative antecedents than affirmative antecedents ($\neg q$:17; q:15; r = 6, N = 14, n.s.).

• **Falsification**

The antecedent predictions are the same as for verification. Since both consequent verification predictions were borne out, the consequent predictions made by falsification were disconfirmed.

At this level of analysis, the level I summary is borne out but in more detail. Matching seems to predominate in the consequent cards and verification in the antecedent cards.

(v) Level III

The level III results for the abstract version are presented in table 7.17. The table shows the result of the relevant comparison and whether the predictions made by each of the three strategies was born out (√) or not (×) (see table 6.7, Chapt 6). Apart form (i) and (iii) these analyses represent the comparative analogues of the Reich & Ruth scores. At this third level of analysis it can be seen that subjects' performance profiles in fact belie the apparent evidence from the lower levels. By just carrying out level I and level II analyses, which is

Table 7.17

Level III results (by rule form) for each of Matching, Verification and Falsification in the Abstract Selection Task.

Prediction	Result	Matching	Verification	Falsification
(i)	$p > \neg p$ (r = 3, N = 19, p < 0.025)	√	√	√
(ii)	$q > \neg q$ (r = 3, N = 16, p < 0.05)	√	√	×
(iii)	$p > \neg p$ (r = 2, N = 18, p < 0.001)	√	√	√
(iv)	$q > \neg q$ (r = 2, N = 16, p < 0.005)	√	×	√
(v)	$\neg p > p$ (r = 6, N = 18, n.s.)	(×	√	√)
(vi)	$q > \neg q$ (r = 4, N = 15, n.s.)	(√	√	×)
(vii)	$\neg p > p$ (r = 2, N = 20, p < 0.0005)	×	√	√
(viii)	$\neg q > q$ (r = 6, N = 13, n.s.)	(×	√	×)

the norm in selection task research, the picture was developing of a mixed strategy of verification on antecedent cards and matching on consequent cards. However, when the predictions made by each strategy for antecedent and consequent cards are separated out by rule type, as in table 7.17 a different picture presents itself. Apart from the *if p*, $\neg q$ rule form all consequent predictions went in the direction predicted by verification. Two out of three of these predictions were also made by matching. None-the-less, these results do reveal that with one exception, even for the consequent cards a verification strategy may best summarise subjects overall performance. A way of further testing this is via the use of Pollard indices.

(vi) Pollard indices

Pollard indices were computed for these results (Pollard & Evans, 1987). The index is computed in a similar way to the matching indices employed in analysing the construction task data. It has only been used on affirmative selection tasks, where for the matching (or verifying index), the score is computed for each subject by assigning scores of +1 for a p or q card selection and a -1 for a $\neg p$ or $\neg q$ card selection. A falsification index can be computed similarly by scoring +1 for a p or $\neg q$ selection and -1 for a $\neg p$ or q card selection. This permits a score ranging between -2 and +2 to be computed for each subject allowing parametric analyses. However, the indices have one limitation, since they are not independent they can not be compared with each other. For a negations paradigm experiment, verification, matching and falsification indices (VI, MI and FI indices respectively) can be computed in similar fashion. Table 7.18 shows the relevant computations. Each column can be totalled to provide global VI, MI and FI indices ranging between +8 and -8. A positive index indicates the presence of the relevant strategy.

VI had a mean of 2.417 and a standard deviation of 2.231. MI had a mean of 1.333 and

Table 7.18

Computing the MI, VI and FI Indices

Rule	MI	VI	FI
(i) If p, q	P,Q - ¬P,¬Q	P,Q - ¬P,¬Q	P,¬Q - ¬P,Q
(ii) If p, ¬q	P,Q - ¬P,¬Q	P,¬Q - ¬P,Q	P,Q - ¬P,¬Q
(iii) If ¬p, q	P,Q - ¬P,¬Q	¬P,Q - P,¬Q	¬P,¬Q - P,Q
(iv) If ¬p, ¬q	P,Q - ¬P,¬Q	¬P,¬Q - P,Q	¬P,Q - P,¬Q

standard deviation of 3.064. FI had a mean 1.667 and a standard deviation of 2.095. A highly significant majority of subjects produced positive indices for both VI (17 +, 1 -, $p <$ 0.0005, Binomial tests) and FI (17 +, 2 -, $p < 0.0005$). However, although the majority of subjects produced positive MIs, this was not significant (12 +, 6 -, $p = 0.119$).

These results confirm the summary of the results obtained at level III. The VI index was higher than the other two indices indicating that verification best summarises subjects performance. A highly significant majority of subjects had positive VIs. However, a highly significant majority also had positive FIs indicating that falsification was also employed. Matching, although in evidence was not adopted by a significant majority of subjects. To test for whether the binary contrast class had any significant effect over standard versions of the task requires a standard abstract baseline result from which to compute the same indices for comparison. It was mentioned above that Reich & Ruth (1982) conducted two versions of the negations paradigm, one a replication of Manktelow & Evans (1979) using low thematic content. In this condition they simply replicated Manktelow & Evans' finding of no facilitation, ie. in this condition the standard abstract result was found. Reich & Ruth (1982) provide the raw data for this condition from which it was possible to compute the Pollard indices for a standard negations paradigm selection task result.

In Reich & Ruth's (1982) results, VI had a mean of 1.082 and a standard deviation of 2.360. MI had a mean of 2.5 and standard deviation of 3.830. FI had a mean 0.417 and a standard deviation of 2.532. A significant majority of subjects produced positive indices for both MI (18 +, 5 -, $p = 0.005$, Binomial tests) and VI (14 +, 4 -, $p = 0.015$). However, for FI the division between positive and negative indices was roughly equal (8 +, 7 -, $p = 0.5$). As would be expected these results reveal a reversal in the numbers of subjects adopting each strategy. Comparing Reich & Ruth's results with the present results reveals the following pattern. Although the MI was lower for the present experiments, this was not significant (independent samples t-test, $t = 1.14$, n.s.). This was largely due to the higher variances found for matching indices than for either FI or VI. Both the FI and the VI were significantly higher in the present experiment than in Reich & Ruth's low thematic content condition (FI: $t = 1.97$, 46 df., $p < 0.05$; VI: $t = 1.83$, 46 df., $p < 0.05$), as predictecd.

The only manipulation which was present in the present experiment and not in Reich & Ruth (1982) or other negations paradigm experiments was the provision of a binary contrast class. Given this simple manipulation subjects are now behaving on the task in the manner predicted by pragmatic context theory. The rule remained disjointly expressed, which means that each card was treated independently. On antecedent cards subjects turn those cards for

which an information gaining eduction is licensed, ie. the TA card. For the consequent cards the task changes to one where subjects are trying to explain what is on the other side. They tend to turn the TC card because the only operative constraint is given by the rule. However, for negative consequent rules, they turn the FC card because the ease of constructing the contrast class in a binary situation evokes the realisation that the FC card is the only card which can be explained by the non-presence (*if* p, $\neg q$) rule form) or presence (*if* $\neg p$, $\neg q$) of p.

Matching predicts that subjects turn the FC card on negative consequent rules because they match the named values. However, this would not predict the crossover observed for TC and FC cards between the two negative consequent rules. For the *if* p, $\neg q$ rule form subjects distribution of selection was TC:4, FC:16, but this was reversed for the *if* $\neg p$, $\neg q$ rule form, TC:13, FC:10. Although it violates the independence assumptions of the test, the results of a Chi-square test indicated a significant crossover ($\chi^2 = 4.52$,, 1 df., $p < 0.05$). If subjects were simply matching named values this would not be predicted. However, if they were constructing the relevant contrast classes in a serial order then this would have been predicted. This will be described in the discussion.

Discussion

In arriving at the conclusion that they should turn the FC card for the *if* p, $\neg q$ rule form, subjects are constructing the contrast class described by the consequent negation. Figure 7.5, indicates the processes involved. The boxes can be thought of as working memory and the content of the boxes as the changing contents of working memory as the various contrast classes are constructed and de-constructed. Take the rule: if blue circle, then not red square. To identify the "not-red square", they construct the contrast class given by the singleton set: blue square. This constructive process focuses attention on red square as the only thing which should not be found with a blue circle. Once the contrast class is constructed then the rule states blue circles are only to be found with blue squares. Given the binary situation this renders obvious that a blue circle must not be on the other side of the red square. However, for the *if* $\neg p$, $\neg q$ rule form, things are slightly more complicated. Subjects begin by constructing the antecedent contrast class, by analogy with above rule they construct red circle. They then construct the consequent contrast class, and realise that red square is the only coloured shape that can be found with the opposite of the contrast class they have just constructed. So a further operation must be performed, to get back to blue circle. The extra processing involved is responsible for the crossover observed in the

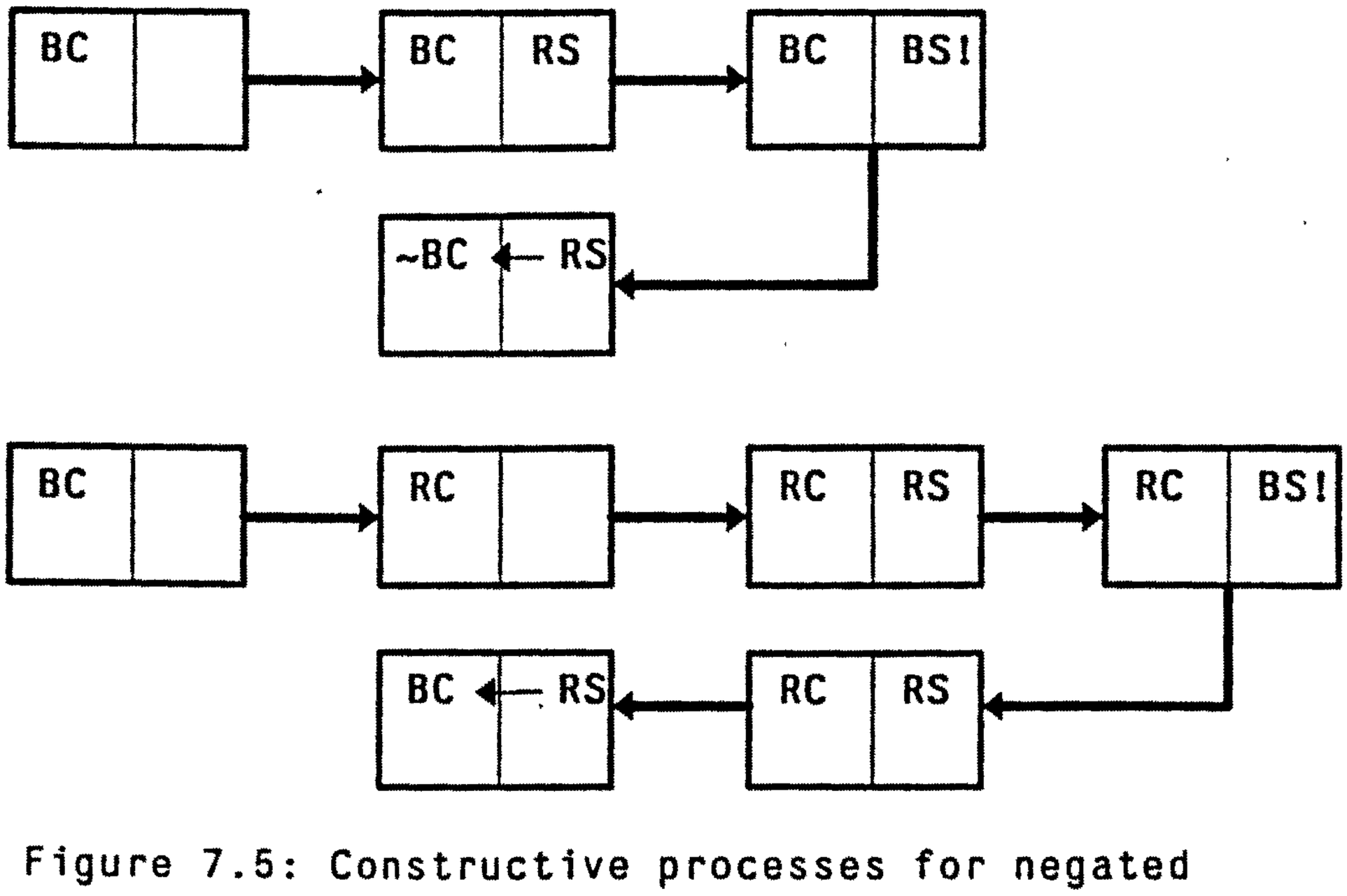

Figure 7.5: Constructive processes for negated consequents, p = Blue Circle
q = Red Square

data.

7.5 Experiment 4: Thematic Selection Task

Introduction

This experiment was the same as Experiment 3 apart from the use of thematic materials. The complete introduction is given in Chapt 6.

Method

Subjects

The same 24 undergraduate psychology students from the University of Edinburgh served as subjects on this experiment as in Experiments 1, 2 and 3. All subjects were tested individually on all conditions.

Design

The design of this experiment was the same as Experiment 3.

Task Materials

The task materials were analogous to those used in Experiment 2, the thematic construction task. Exactly the same rules were used but in the present experiment the antecedent and consequent clauses were either side of the cards. Again this was to maintain homogeneity of lexical content between task types. Whether the boss wants to see me or whether I finish my work on time was stated on one side of each card, and whether I made it home in time for dinner was stated on the other side. All combinations, for the different rules, again yielded the four cards for each rule.

Procedure

The procedure and typed instructions were the same as for Experiment 3, except for the use of the thematic stimulus cards and the following changes in the typed instructions for the first two paragraphs:

(7.11) You will be presented with arrays of four cards like the ones you have just seen and handled, naturally only one side of each card will be visible. On one side of each card something that happens to me at the office is indicated, eg. I finish my work OR the boss doesn't want to see me, and on the other side it is indicated whether or not I make it home for dinner on time.
For each array, you will be given a rule which applies to combinations of the events described on the cards.

Subjects were then told that they could keep the instructions to refer to. The four rules were then presented one at a time on a typed sheet in the order of presentation assigned.

Results

It was argued in Chapt. 6 that when the selection task data is analysed exhaustively there are three levels of comparative analysis. Along with the overall results and a Reich and Ruth (1982) style score, the results will be reported by levels.

(i) Overall Results

Table 7.19 shows the frequency with which subjects gave each logical case as being necessary to test for the truth or falsity of each rule. This table is in the standard format for presenting Evans negations paradigm data. It was argued in Chapt. 6 that only presenting the data by logical case obscured the more detailed levels of analysis which can be performed for verification and falsification when the data are presented by card case. Table 7.20 shows the same data as table 7.19, but by card case.

(ii) Reich & Ruth (1982) Scoring

The results of the same analysis carried out for Experiment 1, were:

(7.12) Verification (25) > Matching (23) > Falsification (19)

Table 7.19

The frequency with which subjects gave each logical case as being necessary to test for the truth or falsity of each rule for the Thematic Selection task.

Rule	TA	FA	TC	FC
(i) If p, q	19	6	14	10
(ii) If p, ¬q	18	7	14	9
(iii) If ¬p, q	16	12	8	9
(iv) If ¬p, ¬q	16	11	11	10
Overall %	77	27	51	43

Table 7.20

The frequency with which subjects gave each card case as being necessary to test for the truth or falsity of each rule for the Abstract Selection task.

Rule	*p*	¬*p*	*q*	¬*q*
(i) If p, q	19	6	14	10
(ii) If p, ¬q	18	7	9	14
(iii) If ¬p, q	12 } 23	16	8	9
(iv) If ¬p, ¬q	11 }	16	10	11
Overall %	55	49	60	33

This result does not accord with Reich and Ruth's (1982), "high thematic content" result:

Thematic Verification (19) > Falsification (15) > Matching (4)

Reasons for this discrepancy are discussed below. This score represents a descriptive analysis for the pattern of results. In the following analyses the data are subject to inferential statistical analysis to further determine the nature of subjects strategies in the present experiment.

(iii) Level I Analyses

- **Matching**

Keeping card case constant, matching predicts (i) more *p* card selections than ¬*p* card selections (*p* > ¬*p*), and (ii) more *q* card selections than ¬*q* card selections (¬*q* > *q*). Prediction (i) was significantly confirmed ($r = 1$, $N = 9$, $p < 0.025$; all results are one tailed sign tests). However, prediction (ii) went in the other direction (*q*:41; ¬*q*:44) albeit not

significantly (r = 4, N = 10, n.s.)

• **Verification**

Keeping logical case constant, verification predicts (i) more TA selections than FA selections (TA > FA), and (ii) more TC selections than FC selections (TC > FC). Prediction (i) was borne out (TA:69; FC:36) but not significantly (r = 6, N =19, n.s.). Similarly for prediction (ii) (r = 5, N = 15, n.s.).

• **Falsification**

Keeping logical case constant, falsification makes the same antecedent prediction as verification, but predicts the exact opposite for the consequent cases, ie. FC > TC. As for verification, prediction (i) was borne out, but prediction (ii) went in the opposite direction. This is expected, since verification and falsification make mutually exclusive predictions for the consequent case and the verification prediction was borne out.

At this level of analysis these results are not as clear cut as for the abstract data. The only significant result was for matching on the antecedent cards. However, both matching and falsification make predictions for the consequent cards which went in the opposite direction to that predicted. Whereas both verification predictions were borne out, albeit not significantly, which seems to tip the balance in favour of this strategy as best summarising subjects' performance.

(iv) Level II

• **Matching**

The four level II matching bias predictions have already been introduced. The prediction that there would be more TA selections for rules with affirmative antecedents than negative antecedents was borne out (p:37; $\neg p$:32), but far from significantly (r = 1, N = 5, n.s.). There were significantly more FA selections for rules with negative antecedents than for those with affirmative antecedents (r = 0, N = 8, p < 0.025). There were *fewer* TC selections for rules with affirmative consequents than negative consequents ($\neg q$:25; q:22) but not significantly. This is in the opposite direction to that predicted by matching. There were equal numbers of FC selections for rules with negative consequents and affirmative

consequents.

• **Verification**

The prediction that there would be more *p* card selections for rules with affirmative antecedents than negative antecedents was confirmed (*p*:37; ¬*p*:32) but not significantly. The prediction that there would be more ¬*p* card selections for rules with negative antecedents than affirmative antecedents was significantly confirmed ($r = 3$, $N = 16$, $p < 0.025$). Although both consequent predictions were borne out neither was significant.

• **Falsification**

The antecedent predictions are the same as for verification. Since both consequent verification predictions were borne out, the consequent predictions made by falsification were disconfirmed.

At this level of analysis, the same picture emerges as for the lower level. However, one of verification's antecedent predictions is now significantly confirmed. Both consequent predictions for both matching and falsification went in the opposite direction.

(v) Level III

The level III results for the abstract version are presented in table 7.21. The table shows the result of the relevant comparison and whether the predictions made by each of the three strategies was born out (√) or not (×) (see table 6.7, Chapt 6). At this more detailed level of analysis the expectation that verification best summarises subjects performance is borne out. It is only at this level that the responses which all strategies predict can be separated out. Despite the lack of significance it can be seen that only 1 (out of 6) of the predictions made by matching went in the predicted direction, as opposed to 5 for verification and 3 for falsification.

Table 7.21

Level III results (by rule form) for each of Matching, Verification and Falsification in the Thematic Selection Task.

Prediction	Result	Matching	Verification	Falsification
(i)	$p > \neg p$ (r = 4, N = 21, p < 0.01)	√	√	√
(ii)	$q > \neg q$ (r = 8, N = 20, n.s.)	√	√	×
(iii)	$p > \neg p$ (r = 4, N = 19, p < 0.02)	√	√	√
(iv)	$\neg q > q$ (r = 7, N = 19, n.s.)	×	√	×
(v)	$\neg p > p$ (r = 7, N = 18, n.s.)	×	√	√
(vi)	$\neg q > q$ (r = 6, N = 13, n.s.)	×	×	√
(vii)	$\neg p > p$ (r = 7, N = 19, n.s.)	×	√	√
(viii)	$\neg q > q$ (r = 8, N = 17, n.s.)	×	√	×

(vi) Pollard indices

Pollard indices were again computed for the thematic version of the selection task. VI had a mean of 1.833 and a standard deviation of 4.14. MI had a mean of 1.333 and standard deviation of 2.091. FI had a mean 1.667 and a standard deviation of 2.427. A significant majority of subjects produced positive indices for both VI (13 +, 4 -, $p = 0.025$, Binomial tests) and FI (12 +, 3 -, $p = 0.018$). However, although more subjects produced positive MIs than negative, this was not significant (7 +, 3 -, $p = 0.119$). Comparing Reich & Ruth's results with the thematic results revealed the following pattern. MI was significantly lower for the thematic version (independent samples t-test, $t = 2.152$, 46 df., $p < 0.025$). Both the FI and the VI were higher in the present thematic experiment than in Reich & Ruth's low thematic content condition, but not significantly (FI: $t = 1.336$, n.s.; VI: $t = 0.756$, n.s.).

In the thematic version matching has failed to account for a significant proportion of the data. These results replicate Reich and Ruth's finding that when appropriate thematic content is available subjects tend to adopt the verificatory strategy licensed by pragmatic context theory. Given the non-taxonomic interpretation and the retention of the inductive structure of the task this was to be expected. The principle difference made by appropriate

thematic content would appear to be a reduction in matching, rather than any, at least significant, increases in verification or falsification.

(vi) Between tasks analysis: selection tasks

In general, the results on both abstract and thematic versions of the selection task revealed similar performance profiles. Verification best summarises subjects behaviour on both tasks. However, there were notable differences between tasks which are tied to subjects' different interpretations of the various rules as revealed in the construction tasks.

No significant or otherwise noteworthy difference in the cards selected between abstract and thematic versions were observed for the *if p, q* rule form. For the *if p, ¬q* rule form, antecedent card selection frequencies were virtually the same. However, there was a significant crossover for the consequent cards. In the abstract version the *q* card was selected significantly more often than the *¬q* card ($p < 0.005$; cf. Table 7.15 (iv)). In the thematic version this was reversed but not significantly (cf. Table 7.18 (iv)). The differences in selection frequencies between tasks for these cards were significant. There were more *q* cards selected in the abstract version than the thematic version ($p < 0.05$, McNemar tests), and there were more *¬q* cards selected in the thematic version than in the abstract version ($p < 0.01$).

For the *if ¬p, q* rule form there were significantly more *q* cards selected in the abstract version than in the thematic version ($p < 0.05$). There were no significant differences in card selection frequencies for the *if ¬p, ¬q* rule form. For the negative antecedent rules in general there were significantly more *p* card selections in the thematic version than in the abstract version ($p < 0.05$).

Discussion

In the abstract version of the selection tasks the binary situation has elicited the eductive behaviour which pragmatic context theory predicts. This is, however, not independent of the disjoint expression of the rule nor the difficulties with processing negations. The disjoint rule expression elicits the predict and explain strategy licensed by the non-taxonomic interpretation. However, the binary situation evokes the realisation that the appearance of *q* can be explained by the non-occurrence of the antecedent (cf. discussion, experiment 3).

Moreover, the failure to observe matching on the antecedent cards also argues strongly that it is the binary situation which has facilitated the appropriate strategies. The crossover between consequent card selections for negative antecedent rules argues against a matching interpretation of this result and for a processing deficit explanation incurred by the serial order in which subjects must construct the antecedent and consequent contrast classes for the *if* $\neg p$, $\neg q$ rule form.

The principle differences between abstract and thematic task performance go in the directions which could have been predicted from pragmatic context theory. For the *if* p, $\neg q$ rule form in the abstract version the binary situation has facilitated the realisation that q should not be found with a p on the other side. This *focusing* of attention is a form of matching but one which is elicited by attending to the semantic content of the rules, not a simple matching of named values. This was not observed for the thematic variant where no appropriate contrast classes are specified, rather the negation simply specifies the non-occurrence of the described event, hence the q events non-occurrence can be explained by the occurrence of the p event.

The significant increase in p card selections for the negative antecedent rules in the thematic variant, combined with overall low levels of selections on the consequent cards, appears to be a function of (i) the XOR/biconditional interpretations elicited in the construction tasks, (ii) an unwillingness on behalf of the subject to turn all the cards. In accordance with pragmatic context theory, although subjects interpret the rule as XOR/equivalence in the construction tasks, this does not mean they should thereby select all the cards. Truth conditions and information conditions are different. However, it would appear that subjects have imported this interpretation into their selection task strategies. The significant increase in p card selections for the thematic version appears to indicate that subjects believe that they can predict from my finishing work to my being home. However, they also believe they can predict from my not finishing work to my not being home. This result is not wholly consistent with pragmatic context theory. But neither are they consistent with matching.

The high number of p card selections also argues against a simple perceptual matching strategy. Since explicit negations were present on the cards in the thematic selection task as well as the thematic construction task, this strategy would have been available to subjects. This would have predicted more verification than matching in the thematic task than the abstract task. However, the results of comparing the Pollard indices for both tasks revealed no significant differences. All the indices were lower for the thematic version but not

significantly (VI: $t = 0.65$, n.s.; MI: $t = 0.913$, n.s.; FI: $t = 0.529$, n.s.). This may be an effect of subjects confusion over the negative antecedent rules in the thematic task. However, perhaps this is the wrong comparison to make, the present abstract selection task results were non-standard, since the binary situation has overridden the strategy adopted when the contrast class is less determinate. However, comparing the thematic results with Reich & Ruth's (1982) low thematic content condition revealed that the principle effect of the thematic materials used here (including the explicit negations) was a significant reduction in matching *without* a correlative increase in verification or falsification levels. This argues against a perceptual matching strategy. The principle effect of the binary situation in the abstract task, contrasts with the effect of thematic materials. Rather than a decrease in matching, significant increases in verification and falsification were observed. This is consistent with the fact that a by-product of constructive negation is the possibility of matching, since subjects must begin the constructive process with the named value (cf. discussion, experiment 3).

Summary

In experiment 1, very few of the predictions made by matching were observed. Instead highly differentiated interpretational effects predominated. The clear distinction between the behaviour observed between affirmative and negative antecedents could not be explained by matching. The only plausible explanations for the anomalous results was either (i) a processing deficit produced by the negations, or (ii) a non-taxonomic rule mis-interpretation. However, experiment 2, discounted (ii). This experiment also demonstrated the predicted interpretational differences between taxonomic and non-taxonomic rules, while simultaneously discounting a simple perceptual matching explanation. A model was provided for the constructive processes of contrast class formation which exploited the concrete presence of the array in the abstract task.

In experiment 3, the use of binary contrast classes was shown to produce a marked facilitation of falsification *and* verification over abstract versions of the negations paradigm selection task. The simultaneous increase in verification was consistent with the fact that the rules were still disjointly expressed, and so the predictive cycle strategy still applicable. Complementary effects were observed for the thematic version in experiment 4. Rather than an increase in falsification and verification a significant decrease in matching was observed. This argued against a simple perceptual matching explanation. However, it appeared that part of this result may be due to subjects importing their XOR/biconditional interpretations

of negative antecedent rules from the construction tasks. Nonetheless, highly differentiated interpretational effects were again observed which argues against matching, even in the selection task.

In the next chapter the consequences of this research for causal modeling of the observed behaviour will be discussed after a summary of the principle conclusions.

Chapter 8: Conclusions and Consequences

8.1 Introduction

In this chapter the main findings of the thesis are summarised by way of a conclusion. In the following sections the consequences of this work are outlined. It will be argued that present theories of conditional reasoning, either based on pragmatic reasoning schema or mental models may be indistinguishable from approaches which assume a mental logic. It is then argued that the most eloquent argument against the doctrine of mental logic is the failure of logical attempts to capture the computation of context. Finally some speculations are offered on the nature of the cognitive architecture which may be able to capture the properties implicit in the conclusions.

8.2 Conclusions

The concept which has played a central role in providing a rational basis for the behaviour observed in human conditional reasoning is the distinction between the logical behaviour of the conditional and its eductive or information gaining behaviour. This contrast has been demonstrated to affect the manner in which people come to fixate conditional beliefs in the model of the predictive cycle. It was shown how this strategy not only accounts for the standard data on the task but also, once the appropriate action orienting interpretation is provided, how the *same* basic strategy is at work in the thematic facilitation results. Any organism needs to be able to predict its environment, and therefore acquire knowledge of the operative constraints in its ecological niche. The mechanisms by which an organism *discovers* those constraints are primarily *bottom-up*. The organism can either discover a constraint by *enumerative* procedures, or *one-shot* processes. Upon regular observation of a correlation between two events, the prior event will be taken as a good predictor of the subsequent event, be this in time or in order of discovery. This process leads to *expectations* which if regularly fulfilled can be only due to the *actual* structure of the world. Only attunements to constraints which are appropriately *grounded* make the planning of *effective* and *efficient* action possible in the long run. Particular salient experiences may also suggest constraints in a *one-shot* manner. Organisms who can be attributed with higher level conscious thought, may also acquire constraints via *top-down* processes. Repeated conscious

decisions can be sedimented into unconscious habits: top-down learning can be enumerative too. But, importantly, promises and moral obligations may provide the basis for attunements which are essentially *one-shot*. The *justificatory* procedure is the same in all cases however, *use* the constraint to predict your world.

Within constraints the distinction between *taxonomic* and *non-taxonomic* constraints proved central. The basis of this distinction is grounded in the contrast between *instances* or single *occurrences* of an event possessing thus and so properties, and *distinct* occurrences being related by some higher order relation. Situation theory allowed the encoding of this distinction, which again has been shown to provide a rational basis for subjects selection task behaviour. In the various COSTs it would appear that this is one of the primary factors responsible for the facilitation of falsificatory responding. The other crucial factor being the *closed domain assumption* which is explicitly incorporated in these tasks. This assumption placed the facilitation observed in these versions of the selection task in a different category to the remaining data. It also questioned the actual falsificatory status of subjects responses, they could have been confirming in a surveyable domain.

By relativising truth to situations, partial bits of the world like the books in Hacking's Borgesian library, situation theory simultaneously embodies the concept of *partial interpretation*. It has been shown how this concept is grounded in the basic context dependence of human reasoning. The constraints or laws which permit effective and efficient action are local. It was demonstrated how this conception of interpretation rendered the strategy of falsification unsound when based either on *modus tollens* or the semantics of the universal quantifier. The domains over which generalisations are stated constantly change in our uncertain world. The concept of partial interpretation was shown to provide a rational basis for the observation of *defective truth* tables in conditional reasoning experiments. And it was demonstrated that the force of this observation has been seriously underestimated in the psychological literature. Experiments were conducted which conclusively demonstrated that subjects behaviour is due to *interpretational* factors and *not* simple response bias, performance error, or shallow linguistic processes.

The observation of defective truth tables occurred in the context of negations paradigm work. It was shown how a *constructive* and essentially *contextually bounded* conception of how negation works which is at least implicit in situation theory, permits a rational foundation to be provided for some of the most recalcitrant data in the psychological literature on conditional reasoning. The ability to construct contrast classes was demonstrated to be influenced not only by peoples pre-existing taxonomic knowledge of the world as they

individuate it, but also crucially on the *circumstantial* features of a subjects environment. The context provided by the card array in the abstract construction class was shown to interact with subjects attempts to construct the relevant contrast classes in a manner which accounts for the systematic errors observed in the data. The circumstantial nature of subjects reasoning with negation is consistent with the conception of the circumstantial nature of human inference articulated in situation theory, (Barwise, 1987).

With these empirically motivated considerations in mind, the two principle psychological theories which could account for subjects reasoning behaviour will looked at. These are *mental models* theory and *pragmatic reasoning schema.* However, other theories are available, for example various schemes of mental logics exist (Braine, Rips etc.). However, it will be argued that the problems which surround these other two theories also infect mental logics.

8.3 Theories of conditional reasoning

Neither mental models nor pragmatic reasoning schema constitute fully articulated theories of human conditional reasoning. No computational models exist capable of deriving similar inference patterns to those observed in conditional reasoning tasks. In this respect, pragmatic reasoning schema may be in better shape than mental models theory. *Prima facie* pragmatic reasoning schema make no appeal to any particular normative theory, in which case there are considerably more degrees of freedom available to extract an empirically valid model. Mental models on the other hand attempt to demonstrate how logically valid inferences are possible, while putting "error" down to performance factors. In Johnson-Laird's model of syllogistic reasoning (Johnson-Laird & Steedman, 1978; Johnson-Laird, 1983) complexity of processing interacts with resource limitations on working memory to produce the observed error patterns.

However, they both share a common characteristic: they may be indistinguishable from theories of mental logic. Let us look at pragmatic reasoning schema first. It was argued in chapter 5, that the production rules provided in Cheng & Holyoak (1985) were in an important sense, vacuous. The inclusion of modal terminology functioned to satisfy intuition but failed completely to provide a satisfactory explanatory account of mechanism. Unless the modal terms were provided with (i) a causal role in the overall production system architecture, or (ii) a denotational semantics, they serve no explanatory role. It was mentioned that the production rules *could* be provided with a semantics along the lines

suggested by Manktelow & Over (personal communication), within the framework of a deontic logic. This render's pragmatic reasoning schema indistinguishable from a mental deontic logic. However, pragmatic reasoning schemas are *modular*. That is they are not part of a monolithic inference regime. They are also context dependent, in the sense that they are accessed not by the particular rule interpretation but by context. However, this does not vitiate against their potentially logical status. It may be that rather than one metanl logic, people possess various mental logics, each accessed to mediate inference in the appropriate contexts.

The appeal to production systems, say as opposed to a logic programming regime as the implementation of the theory, does nothing to distinguish pragmatic reasoning schema from a mental logic. Mental logics have to be implemented too. And the way this is achieved in logic programming is to use a production system. The rule interpreter in PROLOG is a resolution theorem prover which operates over the Horn clause subset of logic (plus Skolemisation to eliminate the quantifiers) which functions as the programming language in which the productions are framed (Hogger, 1984).

Prima facie mental models appear to be distinct from mental logics. Operations are performed over a partial representation of the objects in the appropriate model. The objects are representative arbitrary exemplars of the models domain. However, to implement such a theory requires a notation in which to describe these exemplars. Johnson-Laird argues that it is important to distinguish between syntactic characterisation like mental logics and semantic characterisations like mental models. Johnson-Laird's notation is designed to provide at least a sound system. In the syllogisms model, if the rules which permit the construction of the model and the subsequent operations over it are followed then valid conclusions follow. It is only when the complexity of the derivations this notation employs interacts with resource limitation that error becomes possible. This matching of particular *data structure* and process to the empirical complexity observed in human inference is no mean intellectual feat, but does it license the claim that the notation provided is not a logic? It may be incomplete: would this mean it is not a logic? Well most interesting logics are not complete. Any logic capable of expressing all the truths of Peano arithmetic for example, is not going to be complete (Godel, 1931) - at least if it is complete then it is demonstrably inconsistent (Enderton, 1972). There is a trade off between expressibility and completeness.

There is a fundamental distinction between semantics and syntax. When a language is inductively defined, a recursive function maps those expressions in to the truth values

licensed by the model. Although, the notations which are used to describe both halves of the definition look very much the same it must always be remembered that the semantic half of the definition describes the model directly. Semantic procedures of proof like truth tables or semantic tableaux, operate on semantic principles, unlike purely syntactic procedures like Hilbert or Gentzen style syntactic proof theories. However, they are equivalent. For example, in proving soundness and completeness results for monadic modal logic an equivalence is proved between an axiomatic treatment and semantic tableaux (cf. Hughs & Cresswell, 1968). If people were using one or the other need not be an empirical question. For example, there is no decision procedure for logical implication (|=) for polyadic first order logic. However, by exploiting the provable equivalence of the semantic concept of logical implication (|=) and deducibility (|-), polyadic first order logic is at least recursively enumerable. A decision can be reached in a large but finite amount of time as to the deducibility of a statement but not as to whether it is not deducible. If subjects always drew valid conclusions, then this *a priori* result *could* be taken to indicate that people must be using syntactic procedures.

However, the patent inability of subjecst to draw logically valid inferences in laboratory tasks indicates that this is an empirical question. Would general considerations of complexity enable a decision one way or the other concerning the kind of proof theory implemented, ie. whether it is based on semantic or syntactic principles? Looking to attempts to use logic as a knowledge representation language in Artificial Intelligence is instructive here.

Artificial Intelligence has had a rather cyclical, she loves me, she loves me not, relationship with logic. In the late 60s there were many attempts to use logic in the form of general-purpose theorem provers as general problem solvers (Moore, 1982). A problem situation would be encoded as axioms in first order logic and the to-be-solved problem would be encoded as a theorem which was to be proved from the axioms. This was usually done using the resolution method (Robinson, 1965). However, it was found that, in the general case, the search space grew exponentially (or worse). This meant that only problems of moderate complexity could be solved in a reasonable time. This lead to the view in the AI community that the use of logic as a notation for knowledge representation would be hopelessly inefficient, and therefore should be abandoned. The apparent failure of the programme meant that many of the problems people routinely solve by drawing inferences from their common sense knowledge could not be represented this way and remain tractable. This lead to a move away from logical formalisms during the 70s. This earlier period of AI research engendered an attitude towards logic which is summarised in a quote from

Alan Newell (1980; reported in Moore, 1982):

(8.1) "- the role of logic [is] as a tool for the analysis of knowledge, not for reasoning by intelligent agents"

Newell's statement implies that in analysing a representational/inferential scheme researchers are concerned to ensure that it has the properties guaranteed by a logical formalism. To be a representation of *knowledge* means a formalism must have a *denotational semantics*, it must make sense to ask whether or not the way the world is corresponds to what an expression claims. And indeed AI researchers are keen to provide just such a denotational semantics for whatever formalism they are using, be it, frames, conceptual dependencies, semantic networks etc. Whether or not a standard linear notation is used, if the concern is to provide a denotational semantics for a formalism, then one is engaged in doing logic (Moore, 1982). So, which ever way a researcher turns it would appear that he ends up doing something which looks like logic.

The conclusion reached in this later period of applying logic to problems in Artificial Intelligence, was that more *efficient* ways of implementing logics had to be found: implementations with complexity profiles which did not make such explosive demands on computational resources. There are essentially two ways to achieve this. The first, exploited in PROLOG is to provide a more efficient control regime. The second is to employ different data structures other than the standard linear notation. The question of efficiency is a matter of complexity. Different data structures and different control regimes will produce different complexity profiles. If concern centres on matching the complexity of the implemented logic to the observed human data, then the right control regime and data structure are required. The data structures employed in mental models have been carefully selected to match the empirically observed complexity of drawing certain syllogistic inferences. But this does not imply (i) the notation is not a logic, nor (ii) even if by some other criteria mental models do not constitute a logic, this does not imply that the empirically observed complexity cannot be mimicked by a logic implemented using a novel data structure and/or control regime.

This argument mirrors the representation vs process arguments of the last decade (Anderson, 1978). Perhaps it could have been avoided if in mental models, "error" were not just a function of complexity, but rather of some other more deep seated adaptive function. On pragmatic context theory it is the *computation of context* which is the main determinant of subjects inferential behaviour. How this is achieved, and indeed whether a standard logic

regime can provide a tractable theory of contextualised inference will be looked at next, but not before offering some final comments on pragmatic reasoning schema and mental models.

The data on conditional reasoning and especially Wason's selection task has been taken to argue that human reasoning is content dependent and therefore irrational (eg. Stich, 1985). It has been argued throughout this thesis that (i) in many instances the task is *inductive*, in which case it has been established at least since Goodman (1983, originally, 1954) that content is crucial; (ii) the bulk of the adaptive inferential behaviour of human beings concerns *eductive* inference, which few would deny is context sensitive, and content dependent. Inferential behaviour which is so patently adaptive to the demands of a changing world can hardly be characterised as irrational.

The locus of the problem appears to be with the *form/content* distinction. Once a sufficient grasp on a problem domain has been achieved to the extent that generalisations can be stated it would appear that a formal account will be forthcoming. Both pragmatic reasoning schema and mental models fall foul of this phenomenon. Once the apparently aberrant data is sufficiently organised, essentially formal characterisations are provided which will then be amenable to logical treatment. This is most obvious in the current proliferation of logical systems. Linguistic phenomena like tense, modality etc. are being incorporated into formal logic. The boundary between form and content is apparently flexible. However, will this equally apply to the computation of context? *Prima facie* the observation of contextual effects of the kind observed, described and appealed to in this thesis, are precisely in opposition to the formalising tendency. Observations of contextual effects are precisely observations of the *inability* to state overarching generalisations concerning the inferences subjects should draw. The question that will be addressed in the next section is could even these contextual phenomena be captured logically?

8.4 The computation of context

Within Artificial Intelligence there have been some attempts to deal with contextual phenomena via *default* inference schemes (Reiter, 1980). For example, Johnny believes that all the 1" pipes are threaded but on talking to his boss he knows that it does not hold for bends, it equally may not hold for non-standard lengths of pipe, the new pipes bought in from Japan, copper as opposed to steel pipes etc. All of these possibilities override the generalisation that all the 1" pipes are threaded. If a standard logical implementation is to

be provided all the various conditions that might override Johnny's rule must be explicitly encoded in a default rule and a check performed that none of them apply in any specific case. In the standard approach to default reasoning in knowledge representation the negation as failure procedure achieves this. Each condition is encoded as a negated conjunct in the antecedent. Unfortunately there are sound reasons to suspect that default reasoning schemes along these lines are intractable.

Most standard logical schemes are monotonic. For example, if Johnny infers that the pipe is threaded on discovering it is 1", then his inference is monotone if no additional premise can invalidate his conclusion. In non-monotonic reasoning premises can be added and conclusions *lost*. For example, when Johnny learns that it is also a bend, he will no longer infer it is threaded.

Attempts have been made to extend standard logic to incorporate non-monotonicity. The most notable example is Reiter (1980), who adds a metatheoretic M-operator to first order logic.

(8.2) p & Mq -> q

i.e. given p and the fact that q is not inconsistent with an agents current knowledge, q can be inferred. McDermott (1986) notes that there are two problems with this formulation and the fixed point semantics provided for the M operator. First, Reiter's logic is undecidable and radically intractable in practice. The problem of deciding whether a default rule applies comes down to consistency checking. When Johnny learns that the bends are not threaded he needs to revise his beliefs about pipes and bends. He has ensure that all his other beliefs are consistent with this new piece of knowledge. Consistency checks are a kind to NP hard satisfiability problem (cf. Horowitz & Sahni, 1979). McDermott calls this the "you can't know problem". Second, given the semantics for M the conclusions drawn are usually too weak. Quite often, although p is the conclusion desired, all that follows is p V q, where q is some arbitrary proposition. McDermott calls this the "you don't want to know problem".

This technical problem need not decide against logical approaches to the computation of context. However, there are more general difficulties which beset any logical approach. There are indefinitely many conditions which may override Johnny's inference concerning the threadedness of 1". Indeed they can be invented at will, eg. the pipes are not lead, brass, aluminium, the thread cutting device at the factory has not broken down, the pipe manufacturer has decided its to expensive to supply them ready threaded etc. These counterexamples are not merely rhetorical devices - they would in the right circumstances block

the inference. On a logical approach, it seems that each of these possibilities must be explicitly encoded in the appropriate rule. To avoid an infinite list of default clauses a finite taxonomy must be appealed too which captures the infinitude of specific cases. Perhaps the 1" pipes are unthreaded if they are of non-standard length, they have a foreign origin and so on. However, what counts as a non-standard length is relative to what rule is being considered: 4 feet may be a non-standard length vis vis gas pipes, but not vis a vis water pipes. Foreign origin will vary dependent on the location of Johnny's place of employment. So it seems that the categories in our taxonomy must be rule relative - i.e. the precise sense of 'non-standard length, foreign etc.' must be spelt out in detail *in each rule.* It is a considerable act of faith to believe that such specifications will be forthcoming. Default rules infect lexical inference as well as structural inference.

Intractability problems with default logic suggest that logical attempts to compute context are unpromising. Further attempts could involve a heavy investment in modularity. However, the basis for the modularisation would have to be along the lines of the rule relative taxonomies already discussed, which would make for modularisation on a scale which would ultimately prove vacuous. The level of the module may well have to be the individual rule!

Intractability results concerning logical attempts to compute context argue more eloquently against mental logics than any of the proposals put forward by mental modellers. The observation of contextual effects in human inference *per se* does *not* argue against a logical treatement. It is the practical computational intractability of logical systems which attempt to capture contextual phenomena which legislates against mental logics.

Attempts to provide theories of conditional reasoning are generally under-specified. The existing accounts seem to be characterisable in logical terms and hence fail with regard to the requirement to provide an account of the computation of context. In the next and concluding section of this thesis some speculations will be offered on the kind of general cognitive architecture which may at least cohere with the conclusions derivable for the main body of the thesis presented in section 8.2.

8.5 Implications for the human cognitive architecture

It has already been established that a Classical cognitive architecture based on logical proof theoretic principles (Fodor & Pylyshyn, 1988) is unlikely to be able to compute contextualised inferences. The role of any cognitive architecture is the provision of appropriate data structures and an inference regime. However, an inference regime which works on logical principles would appear an unpromising candidate for the architecture of cognition modulo the empirical data on conditional reasoning. Despite which, the general production system architecture may be appealed to. For example, Anderson's ACT* (1984), seems to embody many of the desirable characteristics indicated in the conclusion section. However, it is also the case that PDP (Rumelhart & McClelland, 1986) systems appear to embody some of these requirements. The strategy of this section will, therefore be to briefly introduce the important features of each system and then to locate some of the various properties stated in the conclusions within these frameworks. The accounts will be highly general and speculative, the intention is to very broadly identify the kind of system which is most compatible with pragmatic context theory.

Anderson's (1983) ACT* production system architecture consists of three basic memory systems: a working memory, a declarative memory and a production memory. Working memory contains the items upon which the system is currently working. It can retrieve from and store information in declarative memory which is a local semantic network, ie. each node corresponds to an individual concept. The locus of control in the system is given by the production memory. If the condition half of the production matches, fully or partially, the contents of working memory, then the action is loaded into working memory. The working memory is the recipient of encoded external input, and the initiator of action. The matching principles for the production memory are governed by three conflict resolution heuristics. Partial matching is one of these, although the system is biased to prefer complete matches.

Prima facie ACT* provides some interesting characteristics which are consistent with many of the requirements implicit in the conclusions. The general concept of the production memory is broadly in line with the concept of the eductive process when attuned to a constraint. Many constraints may be operative at any one time and unlike logical systems where matches need to be perfect, conflict resolution procedures can allow partial matches. Moreover, particularly familiar or salient matches (production strength), or matches which are more specific, will prevail. One major weakness of the conditions in the productions, however, is that they are simply a conjunction of the relevant conditions, in much the same

way as a default rule is conjunction of negated conditions. This implies that similar problems concerning the context dependence of human reasoning may infect ACT*. It is just not possible to encode all the relevant background conditions explicitly in the condition half of the condition/action pair. Another problem concerns the gradual processes involved in production learning. Although, in general, it must be conceded that the processes involved in acquiring an attunement to a constraint are enumerative and bottom-up, they can be one-shot, or top down and one-shot. Subsequent justification may well proceed slowly but it would seem that unless a constraint acquired via a one-shot process is given a very high production strength initially it will always be over taken in the conflict resolution process.

However, further characteristics are desirable. The distinction between taxonomic and non-taxonomic constraints seems to be captured by the distinction between declarative memory and the production memory. Hierarchical taxonomic relations can be seen as part of the declarative memory, which encodes information about the properties of instances or the required constituents of a relation. The production memory embodies information concerning the higher level relations between discrete occurrences of events. Contextualised negation could function as an operator in declarative memory accessing the concepts which are inhibited by the activation of the negated constituent. The principle limitation of ACT* would appear to be a deficiency akin to default logics concerning the ability to handle rich contextual information. Let us now look briefly at PDP systems.

PDP systems consist of a large number of interconnected units or idealised neurons which are massively interconnected. The activation levels on each unit affect the units they are connected too mediated by the weights on the connections. They are essentially distributed in the sense that only patterns of activation over several units correspond to a conceptual unit. They "provide an efficient way of using parallel hardware to implement best-fit searches." (Hinton, McClelland and Rumelhart 1986 p 80)

> (8.3) "One way of thinking about distributed memories is in terms of a very large set of plausible inference rules. Each active unit represents a 'microfeature' of an item, and the connection strengths stand for plausible 'microinferences' between microfeatures. Any particular pattern of activity of the units will satisfy some of the microinferences and violate others. A stable pattern of activity is one that violates the plausible microinferences less than any of the neighbouring patterns." (ibid pp 80-81)

It is contentious as to whether PDP systems constitute rival cognitive architectures or implementational theories (Fodor & Pylyshyn, 1988). My own view is that they are best considered as implementational theories which nonetheless play a central role in explaining the causal antecedents of cognitively mediated behaviour (cf. Chater & Oaksford,

forthcoming). And this is nowhere more so than in the area of how context affects human inferential processes.

Hinton, McClelland and Rumelhart (1986) discuss a novel implementation of semantic nets in PDP hardware (originally in Hinton 1981) which characterises how an implementational theory along PDP lines may capture the context sensitivity of human inference:

> (8.4) "People are good at generalising newly acquired knowledge...If, for example, you learn that chimpanzees like onions you will probably raise your estimate of the probability that gorillas like onions. In a network that uses distributed representations, this kind of generalization is automatic. The new knowledge about chimpanzees is incorporated by modifying some of the connection strengths so as to alter the causal effects of the distributed pattern of activity that represents chimpanzees. The modifications automatically change the causal effects of all similar activity patterns. So if the representation of gorillas is a similar activity pattern over the same set of units, its causal effects will be changed in a similar way." (p 82)

The similarity metric used in automatic generalisation is induced by pattern similarity. This similarity metric need not be specified by the programmer (who antecedently knows which generalisations he wants to go through), but may itself be learnt by the network (Hinton 1987). Notice that gorilla is given 'likes onions' as a default. Yet this default may be overridden by explicitly storing information to the contrary. Further the default rule may be overridden if 'gorilla' has a similar pattern to 'orangutan', whose representation does not include 'likes onions'. The similarity metric gives us plausible inference rules for free, and the autoassociative mechanism weighs up these various soft constraints to settle on the best fit interpretation. Soft constraints just are default rules - and soft constraints are the very fabric of PDP implementations.

Three problems currently beset PDP systems. First, other than the basic dynamics of the settling process, ie, the best fit search, PDP models possess no intrinsic dynamics, ie. no control regime. In ACT* the productions provide the locus of cognitive control, ie. treating each production as an attunement to a constraint, the control regime is what keeps working memory in synch with the world. The productions predict the environment. However, there are attempts to provide PDP systems with a more useful dynamics by encoding temporal dependencies (Elman, 1988). Second, and relatedly, PDP systems seem unable to encode structure. For example, the analytic relations between a relation and it constituents is not mediated by stochastic processes, it is a fact about our categorisation of the world which needs to be structurally embodied in the cognitive system. At the present time, apart from hand partioning the network in to structural modules (Hinton, 1981), PDP systems cannot learn structural relations. Again the encoding of temporal dependencies has been suggested

as a way around this problem. Third, all learning in these systems is enumerative. However, any adequate model of the cognitive system is going to have to allow one shot learning. (PDP systems are not restricted to bottom up learning since it depends on where the inputs are coming from whether a particular system is performing a bottom up enumerative learning procedure or top down).

In conclusion, a combination of the contextualising ability of PDP systems and the dynamical and structural properties of ACT* would form a desirable combination. This suggests the possibility of employing hybrid systems which embody both sets of characteristics in modelling subjects inferential behaviour. Such a hybrid system would, moreover, be at least in the spirit of the dual process theory of Wason & Evans (1975). Contextualising PDP Type 1 processes would be responsible for the integration of knowledge sources which subsequently become available to higher level Type II processes, perhaps providing the default values to the constituents of higher level relations, which may then be subject to further higher level processing if required. If a workable dynamics were specified for PDP systems it could be speculated that only when a prediction fails or becomes otherwise salient will it become consciously available for further processing. Similarly the processes must be reversible, high level conscious decision making must be allowed to affect lower level processes. These possibilities for modelling along with continued testing of the empirical consequences of pragmatic context theory will form the locus of my subsequent research.

References

Akatsuka, N. (1986) Conditionals are discourse bound. Chapter 17 in Traugott, E. C., ter Meulen, A., Reilly, J. S. and Ferguson, C. F. (eds.) *On Conditionals*, pp333-351. Cambridge: Cambridge University Press.

Anderson, J. R. (1978) Arguments concerning representations for mental imagery. *Psychological Review*, 85, 249-277.

Anderson, J. R. (1983) *The Architecture of Cognition.* Cambridge, Mass.: Harvard University Press.

Ayer, A. J. (1972) *Probability and Evidence.* New York: Columbia University Press.

Ayer, A. J. (1980) *Hume.* London: Fontana.

Barwise, J. and Perry, J. (1983) *Situations and Attitudes.* Cambridge, Mass.: MIT Press.

Barwise, J. (1984) The Situation in Logic I. Report No. CSLI-84-2, CSLI, March, 1984.

Barwise, J. (1986) Conditionals and Conditional Information. In Traugott, E. C., Meulen, A., Reilly, J. S. and Ferguson, C. A. (eds.) *On Conditionals.* Cambridge: Cambridge University Press.

Barwise, J. (1987) Lectures on Situation Theory and Situation Semantics.

Barwise, J. (1987) Unburdening the Language of Thought. *Mind and Language*, 2.

Beattie, J. and Baron, J. (1988) Confirmation and Matching Biases in Hypothesis Testing. *Quarterly Journal of Experimental Psychology*, 40A, 269-297.

Belnap, N. D. (1970) Conditional Assertion and Restricted Quantification. *Nous*, 4, 1-15.

Bennett, J. (1971) *Locke, Berkeley, Hume: Central Themes*. Oxford: Oxford University Press.

Bracewell, R. J. and Hidi, S. E. (1974) The solution to an inferential problem as a function of the stimulus materials. *Quarterly Journal of Experimental Psychology*, 26, 351-4.

Braine, M. D. S. (1978) One the relationship between the natural logic of reasoning and standard logic. *Psychological Review*, 85, 1-21.

Bransford, J. D. and Johnson, M. (1972) Contextual prerequisites for understanding:some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, 11, 717-726.

Bransford, J. D., Barclay, J. R. and Franks, J. J. (1972) Sentence memory: a constructive versus interpretive approach. *Cognitive Psychology*, 3, 193-209.

Bransford, J. D. and Johnson, M. K. (1973) Considerations of some problems of comprehension. In Chase, W. G. (ed.) *Visual Information Processing*, pp389-392. New York: Academic Press.

Bransford, J. D. and McCarrell, N. S. (1975) A sketch of a cognitive approach to comprehension: some thoughts about what it means to comprehend. In Weimar, W. B. and Palermo, D. S. (eds.) *Cognition and Symbolic Processes*, pp189-229. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Bree, D. S. (1973) The interpretation of implication. In Elithorn, A. and Jones, D. (eds.) *Artificial and Human Thinking*. Amsterdam: Elsevier Scientific Publications.

Bree, D. S. and G, C. (1976) The difficulty of an implication task. *British Journal of Psychology*, **67**, 579-86.

Bromberger, S. (1965) An Approach to Explanation. In Butler, R. J. (ed.) *Studies in Analytical Philosophy*, pp72-105. BBOX.

Carnap, R. (1923) Uber die Aufgabe der Physick und die Andwendung des Grundsatze der Einfachstheit. *Kant-Studien*, **28**, 90-107.

Carnap, R. (1950) *Logical Foundations of Probability.* Chicago, Ill.: University of Chicago Press.

Cartwright, N. (1983) *How the Laws of Physics Lie.* Cambridge: Cambridge University Press.

Castaneda, H. (1975) *Thinking and Doing*, Volume 7: *The philosophical foundations of institutions.* Dordrecht: D. Reidel.

Chater, N. and Oaksford, M. R. (1988) Autonomy, Implementation and Cognitive Architecture: A Reply to Fodor and Pylyshyn. Submitted to COGNITION.

Cheng, P. and Holyoak, K. (1985) Pragmatic Reasoning Schema. *Cognitive Psychology*, **17**.

Cheng, P. W., Holyoak, K. J., Nisbett, R. E. and Oliver, L. M. (1986) Pragmatic versus syntactic approaches to training deductive reasoning. *Cognitive Psychology*, **18**, 293-328.

Chisolm, R. (1946) The contrary to fact conditional. *Mind*, **55**, 289-307.

Clark, H. H. (1977) Bridging. Chapter 25 in Johnson-Laird, P. N. and Wason, P. C. (eds.) *Thinking: Readings in Cognitive Science*, pp411-420. Cambridge: Cambridge University Press.

Colin, A. J. T. (1980) *Fundamentals of Computer Science.* London: MacMillan Press.

Crain, S. and Steedman, M. J. (1985) On not being led up the garden path: the use of context by the psychological parser. In Dowty, D., Karttunen, L. and Zwicky, A. (eds.) *Natural Language Parsing: Psychological, Computational, and Theoretical perspectives.*

Dennett, D. C. (1978) *Brainstorms: Philosophical Essays on Mind and Psychology.* Montgomery, Vermont: Bradford.

Dretske, F. (1985) Constraints and meaning. *Linguistics and Philosophy*, 8, 9-12.

Dummett, M. A. E. (1978) Realism. In *Truth and other Enigmas.* Cambridge, Mass.: Harvard University Press.

Duncker, K. (1945) On Problem Solving. *Psychological Monographs*, 58.

Elman, J. L. (1988) Finding Structure in Time. CRL Technical Report No. 8801, Center for Research in Language, University of California, San Diego, April, 1988.

Enderton, H. (1972) *A Mathematical Introduction to Logic.* New York: Academic Press.

Evans, J. S. B. T. (1972a) Reasoning with negatives. *British Journal of Psychology*, 63, 213-219.

Evans, J. S. B. T. (1972b) Interpretation and 'matching bias' in a reasoning task. *QJEP*, 24.

Evans, J. S. B. T. and Lynch, J. S. (1973) Matching bias in the selection task. *British Journal of Psychology*, 64, 391-397.

Evans, J. S. B. T. and Wason, P. C. (1976) Rationalisation in a reasoning task. *British Journal of Psychology*, 63, 205-12.

Evans, J. S. B. T. and Newstead, S. E. (1977) Language and reasoning: a study of temporal factors. *Cognition*, 8, 265-283.

Evans, J. S. B. T. (1977) Toward a statistical theory of reasoning. *Quarterly Journal of Experimental Psychology*, 29, 621-35.

Evans, J. S. B. T. (1980a) Current issues in the psychology of reasoning. *British Journal of Psychology*, 71, 227-239.

Evans, J. S. B. T. (1980b) Thinking: experiential and information processing approaches. In Claxton, G. (ed.) *Current Psychological Research: New Directions*. London: Routledge and Kegan Paul.

Evans, J. S. B. T. (1982) *The Psychology of Deductive Reasoning*. London: Routledge and Kegan Paul.

Evans, J. S. B. T. (1983a) Selective processes in reasoning. Chapter 5 in Evans, J. S. B. T. (ed.) *Thinking and Reasoning: Psychological Approaches*, pp135-63. London: Routledge and Kegan Paul.

Evans, J. S. B. T. (1983b) Linguistic determinants of bias in conditional reasoning. *Quarterly Journal of Experimental Psychology*, 35A, 635-644.

Fodor, J. A. (1975) *The Language of Thought*. New York: Thomas Crowell.

Fodor, J. A. (1980) Fixation of belief and concept acquisition. In Piatelli-Palmarini, M. (ed.) *Language and Learning: The Debate between Jean Piaget and Noam Chomsky*. Cambridge, Mass.: Harvard University Press.

Fodor, J. A. (1985) Fodor's guide to mental representation: the intelligent auntie's vade-mecum. *Mind*, 94, 76-100.

Fodor, J. A. (1987) *Psychosemantics: The problem of meaning in the philosophy of mind*. Cambridge, Mass.: MIT Press.

Fodor, J. A. and Pylyshyn, Z. W. (1988) Connectionism and Cognitive Architecture: A Critical Analysis. *Cognition*, 28, 3-71.

Frege, G. (1892) On Sense and Meaning. *Zeitschrift fur Philosophie und philosophische Kritik*, 100, 25-50. Translation by Max Black in Frege (1984) pp157-177.

Gazdar, G. (1979) *Pragmatics, Implicature, Presupposition, and Logical Form.* New York: Academic Press.

Gilhooly, K. J. and Falconer, W. A. (1974) Concrete and abstract terms and relations in testing a rule. *Quarterly Journal of Experimental Psychology*, 26, 355-9.

Godel, K. (1931) Uber Formal Unentscheidbare Satze der Principia Mathematica und Verwandter Systeme, I. *Monatshefte fur Mathematik und Physik*, 38, 173-198.

Goldman, A. (1986) *Epistemology and Cognition.* Cambridge, Mass.: Harvard University Press.

Goodman, N. (1983) *Fact, Fiction and Forecast, 4th Edition.* Cambridge, Mass.: Harvard University Press.

Griggs, R. A. and Cox, J. R. (1982) The elusive thematic-materials effect in Wason's selection task. *British Journal of Psychology*, 73, 407-420.

Haack, S. (1975) *Deviant Logics.* Cambridge: Cambridge University Press.

Hacking, I. (1983) *Representing and Intervening.* Cambridge: Cambridge University Press.

Haiman, J. (1978) Conditionals are topics. *Language*, 54, 564-589.

Hempel, C. (1952) *Fundamentals of Concept Formation in Empirical Science.* Chicago, Ill.: University of Chicago Press.

Hempel, C. (1965) *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: Free Press.

Henle, M. (1962) On the relation between logic and thinking. *Psychological Review*, 69, 366-378.

Hinton, G. E., McClelland, J. L. and Rumelhart, D. E. (1986) Distributed Representations. Chapter 3 in *Parallel Distributed Processing*, Volume 1. Cambridge, Mass.: MIT Press.

Hinton, G. E. (1987) Learning Distributed Representations of Concepts. In *Eighth Annual Conf. of the Cognitive Science Society*, 1987.

Howard, W. A. (1980) The formulae-as-types notion of construction. In Hindley, J. R. and Seldin, J. P. (eds.) *To H. B. Curry, Essays on Combinatory Logic, Lambda Calculus and Formalism*, pp479-490. London: Academic Press.

Hughes, G. E. and Cresswell, M. J. (1968) *An Introduction to Modal Logic*. London: Methuen.

Israel, D. and Perry, J. (1987) What is Information?. Technical Report, CSLI, Stanford University, Stanford, 1987.

Jackson, F. (1985) On the semantics and logic of obligation. *Mind*, 96, 177-96.

Johnson-Laird, P. N. and Tagart, J. (1969) How implication is understood. *American Journal of Psychology*, 2, 367-73.

Johnson-Laird, P. N. and Wason, P. C. (1970) A theoretical analysis of insight into a reasoning task. *Cognitive Psychology*, 1, 134-148.

Johnson-Laird, P. N., Legrenzi, P. and Legrenzi, M. S. (1972) Reasoning and sense of reality. *British Journal of Psychology*, 63, 395-400.

Johnson-Laird, P. N. and Steedman, M. J. (1978) The psychology of syllogisms. *Cognitive Psychology*, **10**, 64-99.

Johnson-Laird, P. N. (1983) *Mental Models*. Cambridge: Cambridge University Press.

Johnson-Laird, P. N. (1986a) Reasoning without Logic. Chapter 1 in Myers, T., Brown, K. and McGonigle, B. (eds.) *Reasoning and Discourse Processes*, pp13-51. London: Academic Press.

Johnson-Laird, P. N. (1986b) Conditionals and mental models. Chapter 3 in Traugott, E. C., ter Meulen, A., Reilly, J. S. and Ferguson, C. A. (eds.) *On Conditionals*, pp55-76. Cambridge: Cambridge University Press.

Klayman, J. and Ha, Y. (1987) Confirmation, Disconfirmation, and Information in Hypothesis Testing. *Psychological Review*, **94**, 211-228.

Kleene, S. C. (1952) *Introduction to Metamathematics*. New York: North Holland.

Koffka, K. (1935) *Principles of Gestalt Psychology*. New York: Harcourt, Brace.

Kripke, S. A. (1965) Semantical analysis of intuitionistic logic I. In Crossley, J. N. and Dummett, M. A. E. (eds.) *Formal Systems and Recursive Functions*. Amsterdam: North Holland.

Kripke, S. (1982) *Wittgenstein on Rules and Private Language*. Cambridge, Mass.: Harvard University Press.

Kuhn, T. (1970) *The Structure of Scientific Revolutions, 2nd Edition*. Chicago, Illinois: The University of Chicago Press.

Lakatos, I. (1970) Falsification and the Methodology of Research Programmes. In Lakatos, I. and Musgrave, A. (eds.) *Criticism and the Growth of Knowledge*, pp91-196. Cambridge: Cambridge University Press.

Lewis, D. K. (1973) *Counterfactuals*. Cambridge, Mass.: Harvard University Press.

Luchins, A. S. (1942) Mechanisation in problem solving: The effect of Einstellung. *Psychological Monographs*, **54**, Whole-issue.

Luchins, A. S. and Luchins, E. H. (1950) New experimental attempts at preventing mechanisation in problem solving. *Journal of General Psychology*, **24**, 326-39.

Manktelow, K. I. and Evans, J. S. B. T. (1979) Facilitation of reasoing by realism: effect or non-effect? *British Journal of Psychology*, **71**, 227-31.

Marslen-Wilson, W. and Tyler, L. K. (1980) The Temporal Structure of Spoken Language Understanding. *Cognition*, **8**, 1-74.

McClelland, J. L. and Rumelhart, D. E. (1980) An Interactive Activation Model of the Effect of Context in Perception. Technical Report No. 91, Centre for Human Information Processing, University of California at San Diego, 1980.

Moore, R. C. (1982) The Role of Logic in Knowledge Representation and Commonsense Reasoning. Technical Note No. 264, SRI International, Menlo Park, Ca., June, 1982.

Neisser, U. (1967) *Cognitive Psychology*. New York: Appleton-Century-Croft.

Nute, D. (1984) Conditional Logics. In Gabbay, D. and Guenter, F. (eds.) *Handbook of Philosophical Logic*, Volume II. Dordrecht: D. Reidel.

Pollard, P. (1979) Human reasoning: logical and non-logical explanations. PhD Thesis, Plymouth Polythechnic.

Pollard, P. (1981) The effect of thematic content on the Wason selection task. *Current Psychological Research*, **1**, 21-30.

Pollard, P. and Evans, J. S. B. T. (1981) The effects of prior belief in reasoning: an associational interpretation. *British Journal of Psychology*, **72**, 73-82.

Pollard, P. and Gubbins, M. (1982) Context and rule manipulations on the Wason selection task. *Current Psychological Research*, 2, 139-150.

Pollard, P. (1982) Human reasoning: Some possible effects of availability. *Cognition*, 12, 65-96.

Pollard, P. and Evans, J. S. B. T. (1987) Content and context effects in reasoning. *American Journal of Psychology*, 100, 41-60.

Popper, S. K. R. (1959) *The Logic of Scientific Discovery*. London: Hutchinson.

Putnam, H. (1974) The 'Corroboration' of Theories. In Schilpp, P. A. (ed.) *The Philosophy of Karl Popper*, Volume I, pp221-240. La Salle, Illinois: The Open Court Publishing Company.

Quine, W. V. O. (1950) *Methods of Logic*. New York: Holt, Rinehart and Winston.

Reich, S. S. and Ruth, P. (1982) Wason's selection task: verification, falsification and matching. *British Journal of Psychology*, 73, 395-405.

Reichenbach, H. (1944) *Philosophical Foundations of Quantum Mechanics*. Berkeley and Los Angeles: University of California Press.

Reiter, R. (1980) A logic for default reasoning. *Artificial Intelligence*, 13, 81-132.

Robinson, J. A. (1965) A Machine-oriented Logic Based on the Resolution Principle. *Journal of the ACM*, 12, 23-41.

Rumelhart, D. E. and McClelland, J. L. (eds.) (1986) *Parallel Distributed Processing: Exploration in the Microstructures of Cognition*, Volume 1: *Foundations*. Cambridge, Mass.: MIT Press.

Rumelhart, D. E., Smolensky, P., McClelland, J. L. and Hinton, G. E. (1986) Schemata and sequential thought processes in PDP models. Chapter 14 in McClelland, J. L. and Rumelhart, D. E. (eds.) *Parallel Distributed Processing: explorations in the microstructure of cognition*, Volume 2: *Psychological and Biological Processes*, pp7-57. Cambridge, Mass.: MIT Press.

Russell, B. (1905) On Denoting. *Mind*, **14**, 479-493.

Ryle, G. (1949) *The Concept of Mind*. London: Hutchinson.

Salmon, W. C. (1971) *Statistical Explanation and Statistical Relevance*. Pittsburgh: University of Pittsburgh Press.

Scott, D. (1971) On Engendering an Illusion of Understanding. *The Journal of Philosophy*, **68**, 787-807.

Seuren, P. (1985) *Discourse Semantics*. London: Routledge and Kegan Paul.

Shoemaker, S. (1980) Causality and Properties. In van Inwagen, P. (ed.) *Time and Cause*, pp109-135. Dordrecht: D. Reidel.

Smalley, N. S. (1974) Evaluating a rule against possible instances. *British Journal of Psychology*, **65**, 293-304.

Stalnaker, R. C. (1968) A theory of conditionals. In Rescher, N. (ed.) *Studies in Logical Theory*. Oxford: Blackwell.

Stalnaker, R. C. (1984) *Inquiry*. Cambridge, Mass.: MIT Press.

Stalnaker, R. C. (1986) Possible worlds and situations. *Journal of Philosophical Logic*, **15**, 109-123.

Stich, S. P. (1985) Could man be an irrational animal?. *Synthese*, **64**, 115-135.

Strawson, P. (1950) On Referring. *Mind*, **59**, 320-44.

Suppe, F. (1977) *The Structure of Scientific Theories*, Volume 2ndEdition. Urbana, Ill.: University of Illinois Press.

Toulmin, S. (1961) *Foresight and Understanding*. London: Hutchinson.

Trusted, J. (1979) *The Logic of Scientific Inference: An Introduction.* London: Macmillan.

van Fraassen, B. C. (1980) *The Scientific Image.* Oxford: Oxford University Press.

Veltman, F. (1985) Logics for Conditionals. PhD Thesis, Faculteit der Wiskunde en Natuurwetenschappen, University of Amsterdam.

Veltman, F. (1986) Data semantics and the pragmatics of indicative conditionals. Chapter 7 in Traugott, E. C., ter Meulen, A., Reilly, J. S. and Ferguson, C. A. (eds.) *On Conditionals*, pp147-168. Cambridge: Cambridge University Press.

von Wright, G. H. (1957) *The Logical Problems of Induction.* BBOX.

Wanner, E. and Maratsos, M. (1978) An ATN Approach to Comprehension. In Halle, M., Bresnan, J. and Miller, G. A. (eds.) *Linguistic Theory and Psychological Reality.* Cambridge, Mass.: MIT Press.

Wason, P. C. (1959) The processing of positive and negative information. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 21, 92-107.

Wason, P. C. (1960) On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, 12, 129-140.

Wason, P. C. (1961) Response to affirmative and negative binary statements. *British Journal of Psychology*, 52, 133-142.

Wason, P. C. (1965) The contexts of plausible denial. *Journal of Verbal Learning and Verbal Behavior*, 4, 7-11.

Wason, P. C. (1966) Reasoning. In Foss, B. (ed.) *New Horizons in Psychology.* Harmondsworth, Middlesex: Penguin.

Wason, P. C. (1969) Regression in reasoning. *British Journal of Psychology*, 60, 471-480.

Wason, P. C. and Johnson-Laird, P. N. (1970) A conflict between selecting and evaluating information in an inferential task. *British Journal of Psychology*, **61**, 509-515.

Wason, P. C. and Shapiro, D. (1971) Natural and Contrived Experience in a Reasoning Problem. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, **23**, 63-71.

Wason, P. N. and Johnson-Laird, P. C. (1972) *Psychology of Reasoning: Structure and Content.* Cambridge, Mass.: Harvard University Press.

Wason, P. C. and Evans, J. S. B. T. (1975) Dual processes in reasoning? *Cognition*, **3**, 141-154.

Wason, P. C. (1980) The verification task and beyond. In Olson, D. R. (ed.) *The Social Foundations of Language and Thought.* New York: Norton.

Wason, P. C. (1983) Realism and rationality in the selection task. Chapter 2 in Evans, J. S. B. T. (ed.) *Thinking and Reasoning: Psychological Approaches*, pp44-75. London: Routledge and Kegan Paul.

Wason, P. C. and Green, D. W. (1984) Reasoning and mental representation. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, **36**, 597-610.

Wertheimer, M. (1985) A gestalt perspective on computer simulations of cognitive processes. *Computers in Human Behaviour*, **1**, 19-33.

Yachanin, S. A. and Tweney, R. D. (1982) The effect of thematic content on cognitive strategies in the four-card problem. *Bulletin of the Psychonomic Society*, **19**, 87-90.